Resource Management in Cloud-based Radio Access Networks: a Distributed Optimization Perspective

A THESIS
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Wei-Cheng Liao

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
Doctor of Philosophy

Zhi-Quan Luo

November, 2015

Acknowledgements

There are many people that I would like to show my appreciation for their precious time spending on me during the last few years. First of all, I would like to express my deepest gratitude toward my advisor, Professor Zhi-Quan Luo, for his continuous support and help during my graduate studies. He has been a wonderful teacher to me. I obtain lots of knowledge in optimization and many other areas from him and the great environment he created, especially, the regular group meetings. Beyond his busy schedule, he invested a lot of time and energy in both individual and group meetings with me. It has been an honor working with him. There have been many times that I have struggled in those research works, and his comments and feedback have always helped me stay on the right track. Moreover, every summer for the last few years, he took great care of me in Xi'an, China, which greatly eased my pressure living in a place without any family member. I am and will be deeply grateful to him for what he has done for me.

I would also like to thank the committee members, Professor Nicholas D. Sidiropoulos, Professor Jarvis D. Haupt and Professor Shuzhong Zhang, for their invaluable comments and suggestions during the preliminary exam and the preparation of this dissertation, which help me improve this work. Furthermore, I would like to thank them for kindly accepting to review my dissertation. Specifically, since my graduate study in my home country, the series of outstanding works on the topic of beamforming design Professor Sidiropoulos did has inspired my research works in the wireless communications. These works also helped me to define some interesting problems in this dissertation. In addition, I would like to thank Professor Haupt for his great teaching, which I learned a lot during my graduate school. These courses also help me to solidify my foundation of the research. Additionally, let me thank Professor Shuzhong Zhang, whose course on

the conic optimization is one of the bases in this dissertation. Actually, the Algorithm 1 in this dissertation is developed from the final project of that course.

I am also grateful to the current and former members of Optimization for Signal Processing and Communication group, specially: Shu Cai, Shu-Hsien Chu, Prof. Quan-Bo Ge, Prof. Mingyi Hong, Prof. Bo Jiang, Mojtaba Kadkhodaie, Prof. Qiang Li, Prof. Jing-Ran Lin, Prof. Ya-feng Liu, Chunxing Ni, Wenqiang Pu, Meisam Razaviyayn, Maziar Sanjabi, Ruoyu Sun, Hung-Wei Tseng, Prof. Xiangfeng Wang, Zhen-Hua Xu, Prof. Zi Xu, Yuan Yuan, and Nan Zhang. In particular, I would like to express me gratitude to Professor Mingyi Hong. He has helped me go through the whole research process such as identifying, defining, and refining the research problems. That provided a different view and improved my ability dealing toward the problems at hand. Additionally, I owe a great deal of gratitude to all my friends who supported me and made my graduate school experience more enjoyable.

Last but not the least, I would like to thank my lovely family. I am deeply indebted to my parents, Kuo-Hsiang Liao and Bao-Tsai Lin, and my sister, Shu-Hui Liao, for their constant encouragement and support. I am also grateful to my beautiful wife, Hsuan-Tzu Wei, without whom this journey would be impossible. Her unconditional love and support helped me overcome the problems.

Dedication

To my wife, Hsuan-Tzu Wei.

Abstract

In this dissertation, we consider the base station (BS) and the resource management problems for the cloud-based radio access network (C-RAN). The main difference of the envisioned future 5G network architecture is the adoption of multi-tier BSs to extend the coverage of the existing cellular BSs. Each of the BS is connected to the multi-hop backhaul network with limited bandwidth. For provisioning the network, the cloud centers have been proposed to serve as the control centers. These differences give rise to many practical challenges. The main focus of this dissertation is the distributed strategy across the cloud centers. First, we show that by jointly optimizing the transceivers and determining the active set of BSs, high system resource utilization can be achieved with only a small number of BSs. In particular, we provide efficient distributed algorithms for such joint optimization problem, under the following two common design criteria: *i)* minimization of the total power consumption at the BSs, and *ii)* maximization of the system spectrum efficiency. In both cases, we introduce a nonsmooth regularizer to facilitate the activation of the most appropriate BSs, and the algorithms are, respectively, developed with Alternating Direction Method of Multipliers (ADMM) and weighted minimum mean square error (WMMSE) algorithm. In the second part, we further explicitly consider the backhaul limitation issues. We propose an efficient algorithm for joint resource allocation across the wireless links and the flow control over the entire network. The algorithm, which maximizes the utility function of the rates among all the transmitted commodities, is based on a decomposition approach leverages both the ADMM and the WMMSE algorithms. This algorithm is shown to be easily parallelizable within cloud centers and converges globally to a stationary solution. Lastly, since ADMM has been popular for solving large-scale distributed convex optimization, we further consider the issues of the network synchronization across the cloud centers. We propose an ADMM-type implementation that can handle a specific form of asynchronism based on the so-called BSUM-M algorithm, a new variant of ADMM. We show that the proposed algorithm converges to the global optimal solution.

Keywords: C-RAN; traffic engineering; resource management; BS management; distributed/parallel algorithm; asynchronous

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The success and proliferation of multimedia rich services and smart mobile devices have fueled the explosive growth in the demand for high speed wireless data service in recent years. To accommodate this demand, service providers have increasingly relied on adding macro/micro/pico/femto base stations (BSs) for better cell coverage and higher level of quality of service (QoS), resulting in a heterogeneous network (HetNet) architecture (see [1] and references therein). Note that, for practical applicability, many of the large number of the deployed BSs is with limited power and/or computation capabilities. The recent LTE-A standard has also advocated this type of architecture for coverage extension [2]. The main strength of this type of cellular network lies in its ability to bring the transmitters and receivers close to each other. Thus, significantly less transmit power is needed to deliver higher signal quality. Moreover, when the intracell and intercell interferences introduced by the close proximity of many transmitters and receivers are properly managed, the system performance can be greatly improved.

In recent years, beyond the traditional orthogonalization techniques, e.g. time division multiple access or frequency division multiple access, interference management has been a major focus of the wireless communication research [3, 4]. In case where the number of deployed BSs is large, however, too much backhaul resources will be required, which is impractical. These BSs also require substantial operational costs in the form of static power consumption for supporting the backhaul connection and the cooling system at the BS sites, complexity for encoding/decoding, etc. [5–9] Therefore,

to keep the operational cost manageable for a more environmental friendly and practical communication scheme, it is necessary to jointly manage interference and selectively activate a subset of BSs while shutting down the rest.

Furthermore, in the next generation radio access networks (RAN), many of the large number of BSs may be connected to the core network without carrier-grade backhaul, e.g., WIFI access points with digital subscriber line [10]. Therefore, the RAN has undergone a major structural change, and a novel RAN management methods must be developed for joint resource allocation in the air interface (e.g., precoder design and scheduling) and traffic engineering within the multi-hop backhaul network (e.g., traffic routing) [11–14].

With the advent of cloud computing technologies, one interesting approach gaining support from both academia and industry, is to manage the entire network by a few cloud centers. The migration of the computation tasks from BSs to a few cloud centers is attractive to the operators, since the computational requirements for each BS can be greatly lowered compared to the traditional ones. This further reduces the costs for deploying the extra BSs, and the network management can be more effective and energy efficient. Such software defined, cloud-based RAN (C-RAN) architecture, see Fig.1.1, has been envisioned as a future 5G architecture, and is expected to achieve 1000x performance improvement over the current 4G technology within the next ten years. However, despite the attractiveness for system performance improvement, the increased heterogeneity, network size, cooperation between BSs, and backhaul capability constraints make interference and resource management for future C-RANs an extremely complex and challenging task. Specifically, in light of the huge network size, it is crucial for the joint design problem to be solvable distributedly and in parallel across cloud centers.

1.1 Literature Review

We briefly summarize the main prior results for BS management, especially on the topics of clustering/activation, and the resource management approaches in C-RAN. Our review consists of two parts. In the first part, the main focus is on the wireless resource management for a RAN without any backhaul network restriction. In the

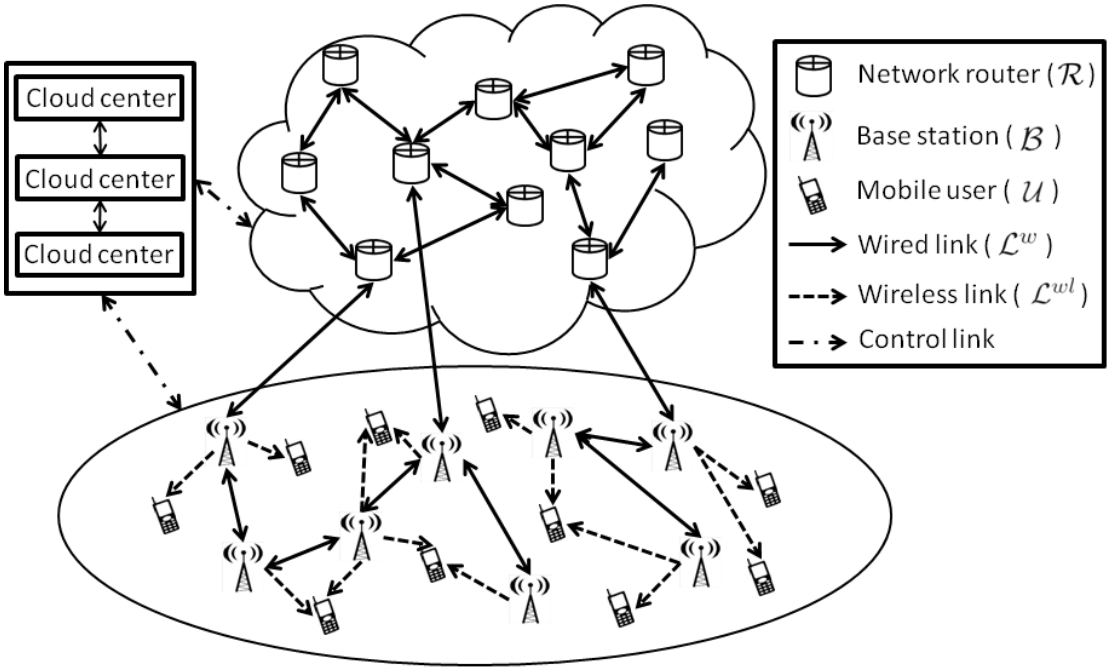


Figure 1.1: Illustration of the considered C-RAN system.

second part, the limitation of the backhaul network for C-RAN is also taken into account. We shall discuss the strength and the limitations of these prior approaches and motivate the research directions we explore in this dissertation.

1.1.1 Prior Results and Approaches on BS Clustering/Activation

In this subsection, we review the recent advances of interference management techniques for the HetNet. The focus will be on the BS clustering/activation techniques aiming to reduce the information exchange overhead and the operational costs for BSs. For the system operators, if the deployed BSs are properly managed, certain system performance metrics, e.g., power consumption, system throughput, or fairness between users, can be greatly improved. These benefits become especially significant when the number of BSs is large. Specifically, in order to effectively manage multiuser interference between BSs, among many existing schemes, e.g., schemes in LTE-A [15], two main modes of BS cooperation have been considered [1]: *(1)* Joint Processing (JP), in which several BSs jointly transmit to users by sharing transmitted data via high speed backhaul network;

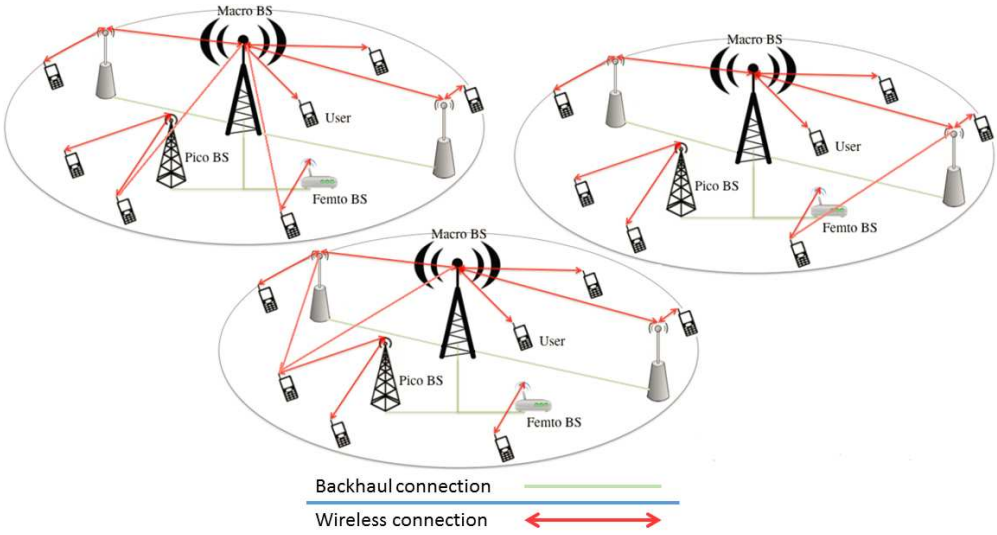


Figure 1.2: Illustration of the downlink multi-cell HetNet.

(2) Coordinated Beamforming (CB), in which BSs mitigate interference by cooperative transmit beamforming without sharing users' data. These two approaches complement each other—JP achieves high spectrum efficiency, while the CB requires less backhaul capacity. In Fig. 1.2, we illustrate the scenario with 3 cells. Within each cell, JP is applied with some high speed back-bone connection while across different cells, CB is applied. For both schemes, the linear beamformers are used at both the transmitters and the receivers such that the interference between the users is mitigated. Since the required backhaul capability differs between them, many recent works propose to strike a balance between the two schemes via BS clustering [16–25]. The idea is to cluster a small number of BSs together such that JP is used only within each BS cluster, and across different clusters, CB is used. Some heuristic approaches that choose the set of BSs according to channel strength [16–18] are insufficient for congestion control, operational cost management, and fairness provisioning. The approaches that design the BS clustering based on the advanced interference management techniques have also been proposed. Along this line of research, two design criteria are mainly used i) minimizing total power consumption and ii) maximizing the utility function.

Particularly, the traditional power minimization beamforming design (see [26–29]) mainly considers the downlink transmission and assumes only one antenna exists at

each mobile user. Mathematically, this can be formulated as

$$\begin{aligned} \min \quad & \sum_{b=1}^{B} P_b \\ \text{s.t.} \quad & \mathrm{SINR}_k \geq \tau_k, \ k = 1, \ldots, K \\ & \text{Other constraints,} \end{aligned}$$

where P_b is the power consumption of BS b and τ_k is the corresponding predetermined desire QoS for user k. We can also consider different practical constraints such as interference being zero-forced between users, i.e., interference alignment and interference neutralization techniques [30–32], the per BS or per antenna power budget, etc. This design problem has mainly been formulated as a second-order cone program (SOCP), which is optimally and efficiently solvable by interior point methods [33] with well-implemented package, e.g., CVX [34]. Furthermore, some distributed approaches [27–29, 35–37] based on uplink-downlink duality or Alternating Direction Method of Multipliers (ADMM) [38, 39] have been proposed. Building on these approaches, recently, [23, 25] propose to add proper sparse regularization terms [40, 41] to the objective function for promoting BS clustering in HetNet. As a result, the trade-off between spectral efficiency and the size of BS clusters is studied. The customized distributed algorithms, which extend the uplink-downlink property to incorporate the extra sparse terms, have also been proposed [23, 25].

To evaluate the effectiveness and fairness of the interference management techniques in terms of the users' data rate, different system utility functions have been proposed. By denoting R_k as the rate of user k, we consider some commonly used utility functions as follow:

- Sum rate utility function: $\mathcal{U}(R_1, \ldots, R_K) = \sum_{k=1}^{K} R_k$
- Harmonic mean utility function: $\mathcal{U}(R_1, \ldots, R_K) = \left(\sum_{k=1}^{K} R_k^{-1} \right)^{-1}$
- Geometric mean utility function: $\mathcal{U}(R_1, \ldots, R_K) = \left(\prod_{k=1}^{K} R_k \right)^{1/K}$
- Min rate utility function: $\mathcal{U}(R_1, \ldots, R_K) = \min_k R_k$

Therefore, the design criterion can be mathematically formulated as

$$\begin{aligned} \max \quad & \mathcal{U}(R_1, \ldots, R_K) \\ \text{s.t.} \quad & \text{constraints as in power minimization design,} \\ & \text{e.g., power budget limitation.} \end{aligned}$$

Such resource management strategies have been extensively studied in the literature not only for the applications of the wireless communications. For example, when there is no signal coding across different antennas, this problem can be reduced to the dynamic spectrum management problem. This is the core problem of digital subscriber line (DSL). Unfortunately, in most well-known utility functions, this problem becomes non-convex and computationally intractable [42]. However, due to its practical importance, different algorithms have been proposed to tackle this problem, e.g., iterative water-filling algorithm (IWFA) [43], autonomous spectrum balancing (ASB) [44], optimal spectrum balancing (OSB) [45], successive convex approximation low complexity (SCALE) algorithm [46], and those in [47–51]. However, these selfish approaches work well only in low interference or signal-to-noise ratio (SNR) cases.

Recently, there have been several significant advances in the utility maximization problem for interference management in wireless communication. Among these, some important results include interference pricing [20, 52, 53], weighted minimum mean square error (WMMSE) algorithm [54–56], multiplicative linear fractional programming-based power allocation (MAPEL) algorithm [57], and partial linearization and convex approximation approach [58]. The detailed theoretical and numerical comparisons between these approaches can be found in [3, 4, 59]. For some special cases, e.g., the min rate utility maximization for interference broadcast channel (IBC) when either the BSs or users are equipped with multiple antennas [26,27,60–62], they have been shown to be polynomial time solvable. However, these techniques are no longer applicable if there is more than one frequency tone or in the presence of backhaul network flow constraints. Moreover, if both BSs and users are equipped with multiple-antennas, the problem is known to be NP-hard [63, 64]. As in the power minimization design criterion, sparse optimization techniques have also been applied to these resource allocation problem to promote the BS clustering [21, 24].

Although the prior schemes on BS clustering for both design criteria have addressed

the trade-off between the effectiveness of interference management and the signaling overhead, none of the them considers the impact of the static operational cost occurred when a large number of BSs are activated simultaneously. As a result, the solutions computed by these algorithms typically require most BSs in the network to remain active. To keep the operational cost manageable, the problem of appropriately selecting a subset of active BSs while shutting down the rest is investigated [22, 65, 66] under the power minimization design criterion. For example, in [22], the BS activation problem is formulated as a mixed-integer optimization, which however, is not practical for large-scale HetNet. In [65], the problem is formulated as a SOCP using sparsity regularizers. In [66], the optimal solution of the BS activation problem is shown to be obtainable by solving an exponential number of semidefinite programming (SDP) problems. Moreover, an effective low complexity heuristic algorithm applying the successive convex approximation technique [46, 58, 67, 68] is proposed. Although SOCP and SDP are convex models which can be solved efficiently, direct optimization using standard package will require a central controller and can be rather slow for a large size network.

1.1.2 Prior Results and Approaches on C-RAN

In this subsection, we review the progress in the resource management and provision of next generation wireless communication system, i.e., C-RAN architecture. This new architecture is characterized by several new features: i) there exist multiple cloud control centers for distributed implementation; and ii) the bandwidth of the backhaul networks is finite. When the number of deployed low power BSs is large, the boundary between each small cell becomes blurred. Moreover, the direct high quality backhaul connection for each of them may not be available. The existing algorithms based on the cellular architecture are therefore not suitable. For C-RAN, the information exchange over the multi-hop backhaul network between these BSs should be explicitly considered.

Specifically, scheduling data traffic for each user within the backhaul network can be viewed as a multi-commodity flow problem. This backhaul traffic engineering problem involves routing from the source nodes (e.g., the cloud centers with backhaul connection) to the destination nodes (e.g., the users requesting content). The resulting optimal solution must guarantee the requested QoS for each end-to-end flow (or commodities in the terminology of traffic engineering) while satisfying the capacity constraints for

all the wireless and/or wired links used by the flows. Compared to the traditional multi-commodity routing in wireline networks [69–73], traffic engineering in the wireless setting is much more challenging. Particularly, the difficulty comes from the multiuser interference of wireless transmission. For each wireless link, the capacity is a nonconvex function of the transmit precoder, and is not known a priori. Moreover, the source-destination pairs depend on the resulting user-BS association, which for C-RAN, is determined by the resulting precoder design. Both are a reflection of a close coupling of the backhaul network and RAN. Therefore, proper and efficient joint provision methods across the backhaul network and RAN for precoding design will be a central component of the newly proposed C-RAN concept, which advocates centralized network provisioning with powerful control centers.

In addition to the BS clustering schemes that implicitly account for the finite backhaul capacity constraints, the impact of the extra backhaul constrains on wireless resource allocation has also been studied recently in the context of joint processing between BSs, e.g., [74–77]. In [74–76], the joint resource allocation, rate adaptation, and user association is investigated. However, the formulated joint problem is too complicated, so certain layering approach has been advocated to improve the system performance without optimality guarantee. In [77], the authors consider the successive interference cancelling (SIC) at the receiver side, and an SDP-based flow rate allocation within the backhaul is proposed. However, these works do not consider the multi-hop routing between the source and the destination nodes.

The joint physical layer precoder design and the backhaul network layer traffic engineering problem for wireless network has also been considered in the framework of cross-layer network utility maximization (NUM) problem, see e.g. [78–85] and some tutorial papers [86–88]. The authors of [78, 87] avoided the nonconvex multiuser interference by considering only the orthogonal wireless links which effectively reduced the problem to convex one. By applying similar approaches, in [80] the multiuser interference is considered but there is no theoretical guarantee of optimality. In [79, 81, 86, 88], the interference was considered in a fast fading environment for which the Lagrange duality gaps can be shown to be zero. But the proposed algorithms required solving difficult subproblems in the dual domain. In [82], the network was approximated by a deterministic channel model [89] through which an approximate optimal solution was

derived. In [83–85], the multiuser interference is handled by exploring the successive convex approximation technique for some special wireless systems that include certain backhaul network routing protocols, e.g., power control with back-pressure [85] and ALOHA medium access control (MAC) [83]. A similar joint optimization problem was also investigated in [90] for a wireless sensor network whereby a distributed algorithm capable of converging to the stationary solution is proposed. However, it requires single antenna nodes and a strongly convex utility function.

Due to the large size of C-RAN architecture, it is crucial that the joint optimization problem can be implemented distributedly and/or in parallel over multiple cloud centers. Most of the existing distributed NUM algorithms is based on the primal/dual decomposition method with subgradient update [78, 87, 91–93]. A closely related framework for the distributed implementation has been proposed by the early works of F. Kelly [72, 73]. Compared with the decomposition methods, this framework is based on some differential equations of the primal variables rather than exploiting duality theory. Through a Lyapunov type analysis on the differential equations, the convergence property can be obtained. The two methods result in similar updating steps while the choice of stepsize for the decomposition method corresponds to the choice of the underlying differential equation for Kelly's framework. However, both approaches can exhibit slow convergence, and they require the optimization problems to be strongly convex. In contrast, a novel parallel and distributed algorithmic framework based on ADMM has been proposed for this C-RAN provisioning problem [37, 94–98]. Empirically, these resulting algorithms are significantly more efficient than the subgradient-based methods, and no strong convexity is needed.

When a large number of networked computation nodes, e.g., BSs and network routers, are coordinating for the updates, network synchronization becomes an important issue for the distributed implementations. On the one hand, perfect synchronization among the network nodes are desirable since the network and the algorithm shall behave in a predictable manner. On the other hand, perfect synchronization results in complicated protocol among the nodes, and the performance of the entire network is determined by the speed of the slowest node. The latter fact is highly relevant in large-scale network processing, as across the network the data sets are often nonuniformly distributed

and the computational power can differ dramatically. Strictly enforcing network synchronization, in this case, scales poorly over large networks [99]. With respect to this synchronization issue, the partially asynchronous [38] version of the dual decomposition again converges slowly, and it further requires that the design objective satisfies some restrictive conditions [92, 100].

Recently, the limited backhaul capability has also been considered when the cloud centers apply the compress-and-forward scheme and JP scheme is used between BSs, e.g., [101–103]. In [101,103], an interesting observation has been made which asserts that under certain scenarios, it may be beneficial to transmit the compressed baseband signals from central unit to each BS instead of the hard information itself. This individual compression of baseband signal for each BS is extended recently to joint compression among BSs [102]. Specifically, it allows the cloud center to compute a joint precoding strategy for all the BSs, and then compress the precoded messages before sending to the BSs via the backhaul; see [102] and the references therein. Here the limited backhaul capacity determines the level of compression needed for each data stream. However, this line of work usually assume that there is a single-hop direct connection between the BSs and the cloud centers (except for [104] with much higher computational complexity), and that the routing of the traffic within the backhaul has been predetermined.

Despite a rich body of literature mentioned above, in the following, we summarize the main challenges arising from the resource allocation and management in cloud-based C-RAN, for which this dissertation tries to handle:

- *Resource Management in RAN*: Most of the existing works on multi-cell interference management assume that the BS-user assignment is known and fixed. However, in C-RAN, users are simultaneously covered by a large number of BSs with different capabilities and loading status. The traditional approach that chooses the set of BSs with strong channel strength is insufficient. Moreover, the finite bandwidth within the backhaul network complicates the choice of BSs. Hence, dynamic selection for the serving BSs becomes a crucial aspect in optimizing the overall system performance.

- *Traffic Engineering Together with Interference Management:* In the context of

C-RAN, the flow rate control for the entire network should be done in conjunction with the multiuser interference management. Hence, the existing works on traffic engineering should be extended to accommodate the nonconvex capacity constraints caused by multiuser interference, which has not been explicitly considered yet. Moreover, even without the backhaul network consideration, most of the formulated problems with multi-user interference are already NP-hard in general [42, 63, 64].

- *Distributed/Parallel Implementation:* Due to the large number of deployed BSs, the size of the resulted joint optimization problem can be huge. As a result, directly solving the problem may still be difficult in real time, even the communication overhead between each agent in the network can be ignored. Up to now, primal/dual decomposition or the framework based on Lyapunov analysis have been applied to exploit the structure of the problem for parallel implementation with easy subproblems. However, the convergence rate strongly depends on whether the problem is *strongly* convex or not. Thus, the computation and communication overheads - caused by the algorithms - adversely affect the overall system throughput. Furthermore, for scalability of the network size, the distributed implementations need to incorporate certain asynchronism among cloud control centers arising from the uneven data distributions and the varying computational capabilities at each node.

1.2 The Main Contributions

In this dissertation, new interference and BS management techniques are proposed to respond the design challenges brought by the huge number of deployed BSs and the C-RAN architecture. Our results show that the BS activation problem for HetNet and the joint interference management and backhaul flow control problem for C-RAN can be efficiently solved in a distributed/parallel way. These approaches are very effective in terms of system performance and the fairness between the users. We further propose a framework of distributed algorithms for large-scale optimization problem that can, to a certain extent, handle the asynchronous issues between cloud control centers. Throughout this research, advanced optimization techniques will be the key to the development

of various practical algorithms. We anticipate that the methodologies developed in this research will significantly influence the design of C-RAN for the next generation communication systems.

Specifically, in Chapter 2, for the BS activation problem, we propose to design optimal downlink transmit beamforming strategies for a HetNet under the following two criteria: **C1)** given a prescribed QoS, minimize the total power consumption, and **C2)** given the power constraints on each BS, maximize the sum rate performance. In contrast to the existing literature on the downlink beamforming design, we impose the additional requirement that these design criteria are met with a small number of BSs. In our formulation, the latter is achieved by imposing certain sparsity patterns in the users' beamformers. Recently, this idea has also been used in different applications in wireless communications, e.g., antenna selection in downlink transmit beamforming [105], joint power and admission control [106], and the joint precoder design with dynamical BS clustering [19–21, 24, 25, 107]. From the complexity standpoint, the problems being considered are computationally challenging: we show that the problem of selecting the minimum number of active BSs that satisfy a given set of QoS constraints is strongly NP-hard for a multi-input single-output (MISO) system. Moreover, the existing works on this joint beamforming design with BS activation scheme [22, 65, 66] have high computational complexity, and no distributed implementation exists. This motivates us to design practical signal processing algorithms for the problems **C1)** and **C2)**. To this end, our contributions are twofold.

- We generalize the traditional power minimization beamforming design by formulating problem **C1)** as a SOCP with an extra group LASSO sparsity regularizer [40]. We develop efficient customized algorithms for **C1)** by exploring the structure of the SOCP and utilizing the ADMM. The main strength of our approach is that each of its steps is simple, closed-form and can be distributed among the BSs. For the special case of power minimization design without the LASSO sparsity regularizer, many distributed approaches have been proposed in the literature, including those based on uplink-downlink duality [27–29, 35] and those based on the ADMM algorithm [36, 96]. Compared to them, our proposed algorithm is computationally more efficient.

- For the design problem **C2)**, we propose a novel single-stage formulation which trades spectrum efficiency with the number of active BSs. Note that, for **C2)**, the sparsity regularizer from **C1)** cannot be trivially applied. Novel modification for the regularization terms is crucial for the effective selection. An efficient algorithm based on the WMMSE algorithm is then devised to compute a stationary solution for the proposed problem. Once again, this algorithm can be solved distributively among different BSs.

The results of this chapter have been previously appeared in

- W.-C. Liao, M. Hong, and Z.-Q. Luo. Base station activation and linear transceiver design for utility maximization in heterogeneous networks. In Proc. of 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 4419-4423, Vancouver, BC, May 2013.

- W.-C. Liao, M. Hong, Y.-F. Liu, and Z.-Q. Luo. Base station activation and linear transceiver design for optimal resource management in heterogeneous networks. IEEE Trans. Signal Process., 62(15):3939-3952, Aug. 2014.

In Chapter 3, we consider the joint flow control and physical layer interference management problem for a large-scale C-RAN. The goal is to maximize the min-rate among all the requested flows. We propose a new algorithm named N-MaxMin extended from the max-min WMMSE algorithm [64] to solve the joint optimization problem. To our knowledge, no existing interference management algorithms can be directly applied to solve the joint network provisioning problem considered herein. For example, [27, 60–62] exploit the structure of signal-to-interference-plus-noise ratio of wireless links for solving min-rate maximization. However, these approaches cannot directly deal with the users' rate, which is essential when the backhaul flow is jointly considered. We further propose a special variable splitting scheme and apply the ADMM method to this problem which allows for efficient distributed and parallel implementation. The contributions of this chapter are summarized below:

- We provide a novel formulation for the joint backhaul traffic engineering and the wireless resource management problem in a C-RAN. A new algorithm named N-MaxMin is proposed to solve the joint optimization problem. This algorithm is

shown to converge to a stationary solution, and can be extended to networks with multi-antenna nodes and/or problems with different utility functions. Furthermore, by solving the formulated problem, a subset of BSs is dynamically selected to serve each user.

- An efficient implementation of the N-MaxMin algorithm is developed by exploiting the problem structure and utilizing the ADMM technique. The resulting algorithm has simple closed-form updates in each step and allows for parallel implementation among cloud centers. These features make the algorithm well suited for the joint provision of backhaul and radio access networks.

The results of this chapter have been previously appeared in

- W.-C. Liao, M. Hong, and Z.-Q. Luo. Max-min network flow and resource allocation for backhaul constrained heterogeneous wireless networks. In Proc. of 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 845-849, Florence, May 2014.

- W.-C. Liao, M. Hong, H. Farmanbar, X. Li, Z.-Q. Luo, and H. Zhang. Min flow rate maximization for software defined radio access networks. IEEE Journal on Selected Areas in Communications, 32(6):1282-1294, Jun. 2014.

- H. Baligh, M. Hong, W.-C. Liao, Z.-Q. Luo, M. Razaviyayn, M. Sanjabi, and R. Sun. Cross-layer provision of future cellular networks: A WMMSE-based approach. IEEE Signal Processing Magazine, 31(6):56-68, Nov. 2014.

- M. Hong, W.-C. Liao, R. Sun and Z.-Q. Luo. Optimization Algorithms for Big Data with Application in Wireless Networks, Big Data Over Networks, Cambridge University Press, 2015.

In Chapter 4, we relax the synchronization requirement for the distributed implementations developed from ADMM approach, e.g., the algorithms in Chapter 2 and Chapter 3. We propose an asynchronous algorithm based on the Block Successive Upper Bound Minimization method of Multipliers (BSUM-M) recently developed in [108]. The latter is a variant of ADMM. In contrast to the existing asynchronous ADMM algorithms [109, 110], the proposed asynchronous algorithm allows each computation node

to have its own local constraints. The feature gives extra design flexibility to distribute the computation loads over each local node. The summary of the contributions in this chapter is as follows.

- We propose an asynchronous BSUM-M algorithm following a *semi-asynchronous* scheme. For this semi-asynchronous scheme, the computation nodes, e.g., cloud control centers, update part of its variables which is coupled to the new incoming information from other nodes. It differs from the well-known *partially asynchronous* scheme described in [38] in the sense that no out-of-sequence communication is allowed.

- The proposed semi-asynchronous scheme is very similar to the randomized version of ADMM [111, 112]. However, in [111, 112], the sequence of variable updates follows a random distribution, which cannot well model the asynchronism due to different speed between nodes. In contrast, the semi-asynchronous scheme studied in this paper is a more practical deterministic counterpart, in which the nodes are updated following an *essentially cyclic* rule [91, 113] while each of them has its own data load and computational capability.

Part of results of this chapter has been previously appeared in

- W.-C. Liao, M. Hong, H. Farmanbar, and Z.-Q. Luo. Semi-asynchronous routing for large scale hierarchical networks. In Proc. of 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 2894-2898, South Brisbane, Apr. 2015.

1.3 Notations

In this dissertation, we will use the following notations. Boldfaced lowercase (resp. uppercase) letters are used to represent vectors (resp. matrices). $\boldsymbol{x}[a]$ is the ath element of the vector $\boldsymbol{x}$, and $\mathbf{A}[a,b]$ denotes the (a,b)th (block) element for matrix $\mathbf{A}$. $vec(\mathbf{A})$ is the vector derived by the vectorization operator on the matrix $\mathbf{A}$, i.e., stacking columns of $\mathbf{A}$ on top of one another. The notation $\mathbf{I}$ denotes the identity matrix, and $\mathbf{0}$ denotes a zero vector or matrix. The superscripts 'H' and 'T', respectively, stand for the conjugate

transpose and transpose. The set of all n-dimensional real and complex vectors are denoted by $\mathbb{R}^n$ and $\mathbb{C}^n$ respectively. The set of all real and complex m-by-n matrices are denoted by $\mathbb{R}^{m\times n}$ and $\mathbb{C}^{m\times n}$, respectively. The indicator function for a set $\mathcal{A}$ is denoted by $1_{x\in\mathcal{A}}$, that is, $1_{x\in\mathcal{A}} = 1$ if $x \in \mathcal{A}$, and $1_{x\in\mathcal{A}} = 0$ otherwise. The projection function to the nonegative orthant is denoted by $(x)^+$, i.e., $(x)^+ \triangleq \max\{0, x\}$. The diagonalization operator and the block diagonalization operator are denoted by $\mathrm{diag}(\cdot)$ and $\mathrm{blkdg}\{\cdot\}$, respectively. Also, the notation $0 \le a \perp b \ge 0$ means that $a, b \ge 0$ and $ab = 0$. Some other notations are summarized in Table 1.1.

Table 1.1: The List of Notations

$\mathcal{V}$	The set of nodes in the network
$\mathcal{R}$	The set of routers
$\mathcal{B}$	The set of BSs
$\mathcal{K}$	The set of cells
$\mathcal{B}_k$	The set of BSs in cell k
$\mathcal{U}$	The set of mobile users
$\mathcal{U}_k$	The set of mobile users in cell k
$\mathbf{v}$	The joint BF vector of the whole network
$\mathbf{v}_{u_k}^{b_k}$	The BF vector from BS b_k to user u_k
$\mathbf{v}^{b_k}$	The BF of BS b_k
$\mathbf{v}_{u_k}$	The joint BF vector for user u_k
$\mathbf{h}_{u_k}^{b_j}$	The channel between BS b_j to user u_k
$\mathbf{h}_{u_k}^{j}$	The channel between cell j to user u_k
F	Number of subchannel tones
M	Number of total commodities in the system
r_m	Data rate for commodity m
d_m	The destination node for commodity m
s_m	The source node for commodity m
$\mathcal{L}$	The set of links
$\mathcal{L}^w$	The set of wired links
$\mathcal{L}^{wl}$	The set of wireless links
$f_{l,m}$	Transmit rate for commodity m on link l
$\mathbf{f}_l$	Transmit rate for all commodities link l
C_l	The capacity for a wired link $l \in \mathcal{L}^w$
v_l	The precoder for wireless link l
d_l	The destination node for link l
s_l	The source node for link l
h_{ln}	The channel between BS s_s to user d_l via subchannel f_l
$I(l)$	The set of interferer to wireless link l
σ_u^2	The noise power at user u
P_b	The power budget at BS b
$\mathcal{N}$	The network $(\mathcal{V}, \mathcal{L})$ with $\mathcal{V}$ being the set of network nodes and $\mathcal{L}$ being the set of directed links
$\mathcal{N}^i$	The ith subnetwork $(\mathcal{V}^i, \mathcal{L}^i)$ with $\mathcal{V}^i \subseteq \mathcal{V}$ and $\mathcal{L}^i \subseteq \mathcal{L}$

Chapter 2

Base Station Activation and Linear Transceiver Design

In a densely deployed HetNet, the number of pico/micro BSs can be comparable with the number of the users. To reduce the operational overhead of the HetNet, proper identification of the set of serving BSs becomes an important design issue. In this chapter, we show that by jointly optimizing the transceivers and determining the active set of BSs, high system resource utilization can be achieved with only a small number of BSs. In particular, we provide formulations and efficient algorithms for such joint optimization problem, under the following two common design criteria: *i)* minimization of the total power consumption at the BSs, and *ii)* maximization of the system spectrum efficiency. In both cases, we introduce a nonsmooth regularizer to facilitate the activation of the most appropriate BSs. We illustrate the efficiency and the efficacy of the proposed algorithms via extensive numerical simulations.

This chapter is organized as follows. In Sec. 2.1, the system model and the problem formulation are introduced for both problems **C1** and **C2**. In Sec. 2.2 and 2.3, detailed algorithms and their analysis are given. In Sec. 2.4, numerical experiments are provided to show the effectiveness of the proposed approaches.

2.1 Signal Model and Problem Statement

Consider a MISO downlink multi-cell HetNet consisting of a set $\mathcal{K} \triangleq \{1, \ldots, K\}$ of cells. Within each cell k, there is a set $\mathcal{B}_k = \{1, \ldots, B_k\}$ distributed BSs which provide service to users located in different areas of the cell. Denote $\mathcal{B} = \bigcup_{k=1}^{K} \mathcal{B}_k$ as the set of all BSs. Assume that in each cell k, a central controller has the knowledge of all the users' data as well as their channel state information (CSI). Its objective is to determine the transmit beamforming vectors for all BSs within the cell. Let $\mathcal{U}_k \triangleq \{1, \ldots, U_k\}$ denote the users located in cell k, and let $\mathcal{U} \triangleq \bigcup_{k=1}^{K} \mathcal{U}_k$ denote the set of all users. Each user $u_k \in \mathcal{U}$ is served jointly by a subset of BSs in $\mathcal{B}_k$. For simplicity of notations, let us assume that each BS has N transmit antennas.

Let us denote $\mathbf{v}_{u_k}^{b_k} \in \mathbb{C}^N$ as the transmit beamformer of BS b_k to user u_k. Define $\mathbf{v} \triangleq \{\mathbf{v}_{u_k}^{b_k} | u_k \in \mathcal{U}_k, b_k \in \mathcal{B}_k, k \in \mathcal{K}\}$ and $\mathbf{v}^{b_k} \triangleq [(\mathbf{v}_{1_k}^{b_k})^H, (\mathbf{v}_{2_k}^{b_k})^H, \ldots, (\mathbf{v}_{U_k}^{b_k})^H]^H$ respectively as the collection of all the beamformers (BF) in the network, and the BFs used by BS b_k. The virtual BF for user u_k, which consists of all the BFs that serve user u_k, is given by $\mathbf{v}_{u_k} \triangleq [(\mathbf{v}_{u_k}^{1_k})^H, (\mathbf{v}_{u_k}^{2_k})^H, \ldots, (\mathbf{v}_{u_k}^{B_k})^H]^H$. Let $s_{u_k} \in \mathbb{C}$ denote the unit variance transmitted data for user u_k, then the transmitted signal of BS b_k can be expressed as

$$\boldsymbol{x}^{b_k} = \sum_{u_k \in \mathcal{U}_k} \mathbf{v}_{u_k}^{b_k} s_{u_k}. \tag{2.1}$$

The corresponding received signal of user u_k is given by

$$\mathbf{y}_{u_k} = \sum_{l \in \mathcal{K}} (\mathbf{h}_{u_k}^{l})^H \mathbf{x}^l + \mathbf{z}_{u_k}, \tag{2.2}$$

where $\mathbf{h}_{u_k}^{b_l} \in \mathbb{C}^N$ denotes the channel vector between the BS b_l to user u_k; $\mathbf{h}_{u_k}^{l} \triangleq \left[(\mathbf{h}_{u_k}^{1_l})^H, \ldots, (\mathbf{h}_{u_k}^{B_l})^H\right]^H \in \mathbb{C}^{NB_l}$ denotes the channel matrix between lth cell to user u_k; $\mathbf{x}^k \in \mathbb{C}^{NB_k}$ is the stacked transmitted signal $[(\mathbf{x}^{1_k})^H, \ldots, (\mathbf{x}^{B_k})^H]^H$ of all BSs in the kth cell; $\mathbf{z}_{u_k} \in \mathbb{C} \sim CN(0, \sigma_{u_k}^2)$ is the additive white Gaussian noise (AWGN) at user u_k. Assuming that each user treats the interference as noise, then the signal-to-interference-and-noise ratio (SINR) measured at the user u_k can be expressed as

$$\mathrm{SINR}_{u_k} = \frac{|\mathbf{v}_{u_k}^H \mathbf{h}_{u_k}^{k}|^2}{\sigma_{u_k}^2 + \sum_{(l,j) \neq (k,u)} |\mathbf{v}_{j_l}^H \mathbf{h}_{u_k}^{l}|^2} \tag{2.3}$$

The achievable rate of user u_k can be expressed as

$$R_{u_k}(\mathbf{v}) = \log\left(1 + \mathrm{SINR}_{u_k}\right).$$

In this work, our objective is to activate a small number of BSs to support efficient utilization of the system resource. Such resource utilization is measured by either one of the following two criteria: **C1)** total transmit power consumption; **C2)** the overall spectrum efficiency. Ignoring the BS activation problem for now, the BF design problem that achieves the minimum power consumption subject to QoS constraint can be formulated as the following SOCP [27]

$$\begin{aligned}
\min_{\{\mathbf{v}^{b_k}\}} \ & \sum_{b_k\in\mathcal{B}} \|\mathbf{v}^{b_k}\|_2^2 \\
\text{s.t. } & \|\mathbf{v}^{b_k}\|_2^2 \le P_{b_k}, \quad \forall\ b_k \in \mathcal{B} \\
& |\mathbf{v}_{u_k}^H \mathbf{h}_{u_k}^k| \ge \sqrt{\tau_{u_k}\left(\sigma_{u_k}^2 + \sum_{(l,j)\neq(k,u)} |\mathbf{v}_{j_l}^H \mathbf{h}_{u_k}^l|^2\right)}, \\
& \mathrm{Im}(\mathbf{v}_{u_k}^H \mathbf{h}_{u_k}^k) = 0, \ \forall\ u_k \in \mathcal{U},
\end{aligned} \tag{2.4}$$

where τ_{u_k} is the prescribed minimum SINR level for user u_k; P_{b_k} is the power budget of BS b_k, $\forall b_k \in \mathcal{B}$, and Im denotes the imaginary part of a complex number. It turns out that this problem is convex thus can be solved to global optimality [27] in polynomial time.

A related BF design problem that achieves the maximum spectrum efficiency can be formulated as the following sum rate maximization problem

$$\begin{aligned}
\max_{\mathbf{v}} \ & \sum_{k\in\mathcal{K}} \sum_{u_k\in\mathcal{U}_k} R_{u_k}(\mathbf{v}) \\
\text{s.t. } & (\mathbf{v}^{b_k})^H \mathbf{v}^{b_k} \le P_{b_k}, \quad \forall b_k \in \mathcal{B}.
\end{aligned} \tag{2.5}$$

Unfortunately, it is well-known that problem (2.5) is strongly NP-hard in general, thus it is not possible to compute its global optimal solution in polynomial time [42].

In the following sections, we will generalize problems (2.4) and (2.5) by incorporating nonsmooth sparsity regularizers for BS activation, and then develop algorithms that can effectively solve the new formulations.

2.2 Base Station Activation for Power Minimization

2.2.1 The Complexity for BS Activation

Suppose all the BSs are activated, then finding the minimum transmit power that satisfies a given QoS requirement can be formulated in (2.4). We are interested in further requiring that the QoS targets are supported by *the minimum number of BSs*. A natural two-stage approach is to first find the smallest set of BSs that can support the QoS requirements, followed by solving problem (2.4) using the set of selected BSs. In particular, a specific BS can be shut down when its BF is a zero vector, i.e., no transmission occurred. As an illustrating example, in Fig. 2.1, BS 4_k is viewed as being turned down since all subvectors of BF for BS 4_k are zero vectors. Therefore, to support the QoS targets, finding the minimum number of BSs is transformed to finding the BF such that there are as many BSs with zero vector as possible. Applying this observation, the first stage problem is given by

$$\begin{aligned} \min_{\{\mathbf{v}^{b_k}\}} \ & \|\{\|\mathbf{v}^{b_k}\|_2\}_{b_k\in\mathcal{B}}\|_0 \\ \text{s.t.} \ & \|\mathbf{v}^{b_k}\|_2^2 \le P_{b_k}, \quad \forall\ b_k \in \mathcal{B} \\ & |\mathbf{v}_{u_k}^H \mathbf{h}_{u_k}^k| \ge \sqrt{\tau_{u_k}\left(\sigma_{u_k}^2 + \sum_{(l,j)\neq(k,u)} |\mathbf{v}_{jl}^H \mathbf{h}_{u_k}^l|^2\right)}, \\ & \mathrm{Im}(\mathbf{v}_{u_k}^H \mathbf{h}_{u_k}^k) = 0, \quad \forall\ u_k \in \mathcal{U} \end{aligned} \tag{2.6}$$

where the ℓ_0-norm $\|\boldsymbol{x}\|_0$ denotes the number of nonzeros elements in a vector $\boldsymbol{x}$.

It turns out that this two-stage approach can be reformulated into a *single-stage problem* shown below

$$\begin{aligned} \min_{\{\mathbf{v}^{b_k}\}} \ & \|\{\|\mathbf{v}^{b_k}\|_2\}_{b_k\in\mathcal{B}}\|_0 + \theta \sum_{b_k\in\mathcal{B}} \|\mathbf{v}^{b_k}\|_2^2 \\ \text{s.t.} \ & \|\mathbf{v}^{b_k}\|_2^2 \le P_{b_k}, \quad \forall\ b_k \in \mathcal{B} \\ & |\mathbf{v}_{u_k}^H \mathbf{h}_{u_k}^k| \ge \sqrt{\tau_{u_k}\left(\sigma_{u_k}^2 + \sum_{(l,j)\neq(k,u)} |\mathbf{v}_{jl}^H \mathbf{h}_{u_k}^l|^2\right)}, \\ & \mathrm{Im}(\mathbf{v}_{u_k}^H \mathbf{h}_{u_k}^k) = 0, \ \forall\ u_k \in \mathcal{U}, \end{aligned} \tag{2.7}$$

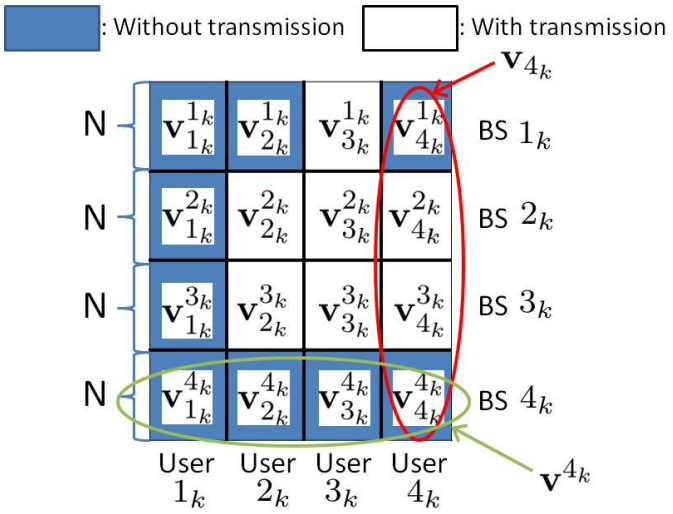


Figure 2.1: Illustration of the sparse precoder pattern at cell k. In this example, BS 4_k is shut down due to no transmission occurred.

where $\theta := \frac{1}{\sum_{b_k \in \mathcal{B}} P_{b_k}}$. The following lemma establishes the relationship among problem (2.7), (2.6) and (2.4).

Lemma 1 *The optimal objective value of problem* (2.7) *lies in* $[S,\ S+1)$ *if and only if the optimal objective value of problem* (2.6) *is* S. *Furthermore, among all solutions with the optimal active BS size equal to* S, *solving problem* (2.7) *gives the minimum power solution.*

Proof Suppose $\mathbf{v}^\star$ is an optimal solution of problem (2.6), which yields the optimal objective value S. Then the objective value of problem (2.7) is $S + \frac{1}{\sum_{b_k \in \mathcal{B}} P_{b_k}} \sum_{b_k \in \mathcal{B}} \|\mathbf{v}^{b_k,\star}\|_2^2 \in [S,\ S+1)$. On the other hand, suppose $\mathbf{v}$ is optimal for problem (2.7) that achieves an objective within the interval $[S,\ S+1)$. Then the optimal solution for (2.6) cannot be smaller than S. Suppose the contrary, that $\mathbf{v}^\star$ satisfies $\|\{\|\mathbf{v}^{b_k\star}\|_2\}_{b_k \in \mathcal{B}}\|_0 \leq S-1$. Then we have

$$\|\{\|\mathbf{v}^{b_k\star}\|_2\}_{b_k \in \mathcal{B}}\|_0 + \theta \sum_{b_k} \|\mathbf{v}^{b_k\star}\|_2^2 \leq -1 + S + \theta \sum_{b_k} \|\mathbf{v}^{b_k\star}\|_2^2 < S,$$

which contradicts the optimality of $\mathbf{v}$. The last claim is also easy to see by a contradiction argument. □

Unfortunately, despite the fact that solving the power minimization problem (2.4) is easy, finding the minimum power *and* the minimum number of BSs for a given set of QoS targets turns out to be difficult. The following result makes this claim precise. We refer the readers to Appendix B for the proof.

Theorem 1 *Solving problem* (2.7) *is strongly NP-hard in the number of BSs, for all* $N \geq 1$.

Motivated by the above NP-hardness result, we proceed to design low-complexity algorithms that can obtain high-quality solutions for problem (2.7). To this end, we propose to use a popular relaxation scheme for this type of ℓ_0-norm minimization problems (e.g., [40]), which replaces the nonconvex ℓ_0-norm by the ℓ_1-norm. Specifically, $\|\{\|\mathbf{v}^{b_k\star}\|_2\}_{b_k\in\mathcal{B}}\|_0$ is relaxed to $\|\{\|\mathbf{v}^{b_k\star}\|_2\}_{b_k\in\mathcal{B}}\|_1 = \sum_{b_k\in\mathcal{B}}\|\mathbf{v}^{b_k}\|_2$. Hence, the relaxed version of the single-stage problem (2.7) can be expressed as

$$f^{\min}(\mathbf{v}) = \min_{\{\mathbf{v}^{b_k}\}} \sum_{b_k\in\mathcal{B}} \beta_{b_k}\|\mathbf{v}^{b_k}\|_2 + \theta \sum_{b_k\in\mathcal{B}} \|\mathbf{v}^{b_k}\|_2^2 \tag{2.8a}$$

$$\text{s.t. } \|\mathbf{v}^{b_k}\|_2^2 \leq P_{b_k}, \ \ \forall\ b_k \in \mathcal{B} \tag{2.8b}$$

$$|\mathbf{v}_{u_k}^H \mathbf{h}_{u_k}^k| \geq \sqrt{\tau_{u_k}\left(\sigma_{u_k}^2 + \sum_{(l,j)\neq(k,u)} |\mathbf{v}_{jl}^H \mathbf{h}_{u_k}^l|^2\right)}, \tag{2.8c}$$

$$\mathrm{Im}(\mathbf{v}_{u_k}^H \mathbf{h}_{u_k}^k) = 0, \ \forall\ u_k \in \mathcal{U}, \tag{2.8d}$$

where $\beta_{b_k} \in \mathbb{R}, \ \ \forall b_k \in \mathcal{B}$ are additional given parameters to further control the number of active BSs of the obtained solution of problem (2.8), c.f. Sec. 2.2.4. In Sec. 2.2.4, we will further discuss how these parameters can be adaptively chosen. Since problem (2.8) is a SOCP (just like problem (2.4)), it can be solved to global optimality using a standard package such as CVX [34]. However, using general purpose solvers can be slow, especially when the number of variables $\sum_{k\in\mathcal{K}} NB_kI_k$ and the number of constraints $2|\mathcal{U}| + |\mathcal{B}|$ become large.

In what follows, we will exploit the structure of the problem at hand, and develop a fast distributed algorithm for solving problem (2.8). Our approach is based on the well-known ADMM algorithm [39], which we outline briefly below.

2.2.2 The Proposed ADMM Approach

In this subsection, we will show that our joint BS activation and power minimization problem (2.8) can be in fact solved very efficiently by using the ADMM, which is briefly reviewed in Appendix A.

The main idea is to decompose the tightly coupled network problem into several subproblems of much smaller sizes, each of which can be solved in closed form. For example, by introducing a copy $\boldsymbol{w}^{q_k}$ for the original BF $\mathbf{v}^{q_k}$, the objective function of problem (2.8) can be separated into two parts

$$\sum_{b_k \in \mathcal{B}} \beta_{b_k} \|\boldsymbol{w}^{b_k}\|_2 + \theta \sum_{b_k \in \mathcal{B}} \|\mathbf{v}^{b_k}\|_2^2, \tag{2.9}$$

where each part is further separable among the BSs. In this way, after some further manipulation which will be shown shortly, it turns out that solving the subproblem for either $\boldsymbol{w}$ or $\mathbf{v}$ can be made very easy.

Formally, let us introduce a few new variables

$$K_{u_k}^{j_l} = (\mathbf{h}_{u_k}^l)^H \mathbf{v}_{j_l}, \ \forall \ u_k, j_l \in \mathcal{U}, \tag{2.10a}$$

$$\boldsymbol{w}^{b_k} = \mathbf{v}^{b_k}, \ \forall \ b_k \in \mathcal{B}, \tag{2.10b}$$

$$\kappa_{u_k} = \hat{\kappa}_{u_k} = \sigma_{u_k} \in \mathbb{R}, \ \forall \ u_k \in \mathcal{U}. \tag{2.10c}$$

and define $\mathbf{K} \triangleq \{K_{u_k}^{j_l} \mid u_k, j_l \in \mathcal{U}\}$, $\boldsymbol{w} \triangleq \{\boldsymbol{w}^{b_k} \mid b_k \in \mathcal{B}\}$, $\mathbf{v} \triangleq \{\mathbf{v}^{b_k} \mid b_k \in \mathcal{B}\}$, $\boldsymbol{\kappa} \triangleq \{\kappa_{u_k} \mid u_k \in \mathcal{U}\}$ and $\hat{\boldsymbol{\kappa}} \triangleq \{\hat{\kappa}_{u_k} \mid u_k \in \mathcal{U}\}$. Clearly $K_{u_k}^{j_l}$ represents the interference level experienced at user u_k contributed by the BF for user j_l; $\boldsymbol{w}^{b_k}$ is a copy of the original BF $\mathbf{v}^{b_k}$; κ_{u_k} and $\hat{\kappa}_{u_k}$ are copies of the noise power σ_{u_k}.

With these new variables, problem (2.8) can be equivalently expressed as

$$\min_{\{\mathbf{v}^{b_k}\},\{\boldsymbol{w}^{b_k}\},\{K_{u_k}^{j_l}\},\{\kappa_{u_k}\},\{\hat{\kappa}_{u_k}\}} \ \sum_{b_k \in \mathcal{B}} \beta_{b_k} \|\boldsymbol{w}^{b_k}\|_2 + \theta \sum_{b_k \in \mathcal{B}} \|\mathbf{v}^{b_k}\|_2^2 \tag{2.11a}$$

$$\text{s.t.} \ \|\boldsymbol{w}^{b_k}\|_2^2 \le P_{b_k}, \ \ \forall \ b_k \in \mathcal{B} \tag{2.11b}$$

$$|K_{u_k}^{u_k}| \ge \sqrt{\tau_{u_k} \left(\kappa_{u_k}^2 + \sum_{(l,j) \ne (k,u)} |K_{u_k}^{j_l}|^2 \right)}, \tag{2.11c}$$

$$\text{Im}(K_{u_k}^{u_k}) = 0, \ \forall \ u_k \in \mathcal{U}, \tag{2.11d}$$

$$(2.10a), \ (2.10b), \ \text{and} \ (2.10c). \tag{2.11e}$$

The partial augmented Lagrangian function of the above problem is given by (2.12),

$$
\begin{aligned}
L(\boldsymbol{w},\mathbf{K},\boldsymbol{\kappa},\mathbf{v},\hat{\boldsymbol{\kappa}},\boldsymbol{\mu},\boldsymbol{\lambda},\boldsymbol{\delta}) &= \sum_{b_k\in\mathcal{B}}\beta_{b_k}\|\boldsymbol{w}^{b_k}\|_2+\theta\sum_{b_k\in\mathcal{B}}\|\mathbf{v}^{b_k}\|_2^2+\sum_{u_k\in\mathcal{U}}(\kappa_{u_k}-\hat{\kappa}_{u_k})\delta_{u_k}\\
&+\mathrm{Re}\left(\sum_{u_k,j_l\in\mathcal{U}}\langle K_{u_k}^{j_l}-(\mathbf{h}_{u_k}^{l})^H\mathbf{v}_{j_l},\mu_{u_k}^{j_l}\rangle\right)+\mathrm{Re}\left(\sum_{b_k\in\mathcal{B}}\langle\boldsymbol{w}^{b_k}-\mathbf{v}^{b_k},\boldsymbol{\lambda}^{b_k}\rangle\right)\\
&+\frac{\rho}{2}\sum_{u_k,j_l\in\mathcal{U}}\left|K_{u_k}^{j_l}-(\mathbf{h}_{u_k}^{l})^H\mathbf{v}_{j_l}\right|^2+\frac{\rho}{2}\sum_{b_k\in\mathcal{B}}\|\boldsymbol{w}^{b_k}-\mathbf{v}^{b_k}\|_2^2+\frac{\rho}{2}\sum_{u_k\in\mathcal{U}}(\kappa_{u_k}-\hat{\kappa}_{u_k})^2,
\end{aligned}
\tag{2.12}
$$

where $\boldsymbol{\mu}\triangleq\{\mu_{u_k}^{j_l}\in\mathbb{C}\mid u_k,j_l\in\mathcal{U}\}$, $\boldsymbol{\lambda}\triangleq\{\boldsymbol{\lambda}^{b_k}\in\mathbb{C}^{I_k}\mid b_k\in\mathcal{B}\}$, and $\boldsymbol{\delta}\triangleq\{\delta_{u_k}\in\mathbb{R}\mid u_k\in\mathcal{U}\}$ are, respectively, the Lagrangian dual variables for the constraints (2.10a), (2.10b), and (2.10c).

It can be readily observed that problem (2.11) is separable among the block variables $\{\mathbf{v},\hat{\boldsymbol{\kappa}}\}$ and $\{\boldsymbol{w},\mathbf{K},\boldsymbol{\kappa}\}$. Specifically, besides the consistency constraints (2.10), the rest of the constraints only depend on variables $\{\boldsymbol{w},\mathbf{K},\boldsymbol{\kappa}\}$. Moreover, it is easily observed that each term in the objective only depends a single block variable. Furthermore, all the constraints linking these two block of variables (i.e., (2.10)) are linear equalities. Therefore, the ADMM algorithm can be directly applied to solve problem (2.11). The main algorithmic steps are summarized in Algorithm 1, which is summarized in Table 2.1.

Before further investigating how each update procedure can be solved in closed-form, let us first discuss the convergence result for the proposed algorithm.

Theorem 2 *Assume that problem* (2.8) *is feasible. Every limit point* $\mathbf{v}^{(t)}$ *(or* $\boldsymbol{w}^{(t)}$*) generated by Algorithm 1 is an optimal solution of problem* (2.8)*.*

Proof. Let us stack all elements of $\{\boldsymbol{w},\mathbf{K},\boldsymbol{\kappa}\}$ and $\{\mathbf{v},\hat{\boldsymbol{\kappa}}\}$ to vectors $\{\boldsymbol{w}_{stack}\in\mathbb{C}^{N|\mathcal{B}||\mathcal{U}|},\mathbf{K}_{stack}\in\mathbb{C}^{|\mathcal{U}|^2},\boldsymbol{\kappa}_{stack}\in\mathbb{R}^{|\mathcal{U}|}\}$ and $\{\mathbf{v}_{stack}\in\mathbb{C}^{N|\mathcal{B}||\mathcal{U}|},\hat{\boldsymbol{\kappa}}_{stack}\in\mathbb{R}^{|\mathcal{U}|}\}$. Then, by comparing problem (A.1) and problem (2.11), when $\boldsymbol{x}=[\boldsymbol{w}_{stack}^H,\mathbf{K}_{stack}^H,\boldsymbol{\kappa}_{stack}^H]^H$

Algorithm 1: ADMM for (2.8):

1: **Initialize** all primal variables $\boldsymbol{w}^{(0)}, \mathbf{v}^{(0)}, \mathbf{K}^{(0)}$ (do not need to be feasible for problem (2.11)); Initialize all dual variables $\boldsymbol{\mu}^{(0)}, \boldsymbol{\lambda}^{(0)}$; $t = 0$;

2: **Repeat**

3: Solve the following problem and obtain $\{\boldsymbol{w}^{(t+1)}, \mathbf{K}^{(t+1)}, \boldsymbol{\kappa}^{(t+1)}\}$ ((2.14), (2.20))

$$\begin{aligned} \min_{\boldsymbol{w},\mathbf{K},\boldsymbol{\kappa}} \; & L(\boldsymbol{w}, \mathbf{K}, \boldsymbol{\kappa}, \mathbf{v}^{(t)}, \hat{\boldsymbol{\kappa}}^{(t)}, \boldsymbol{\mu}^{(t)}, \boldsymbol{\lambda}^{(t)}, \boldsymbol{\delta}^{(t)}) \\ \text{s.t. } & \|\boldsymbol{w}^{b_k}\|_2^2 \le P_{b_k}, \quad \forall b_k \in \mathcal{B} \\ & |K_{u_k}^{u_k}| \ge \sqrt{\tau_{u_k}\left(\kappa_{u_k}^2 + \sum_{(l,j)\neq(k,u)} |K_{u_k}^{j_l}|^2\right)}, \\ & \mathrm{Im}(K_{u_k}^{u_k}) = 0, \; \forall \; u_k \in \mathcal{U}; \end{aligned}$$

4: Solve the following problem and obtain $\mathbf{v}^{(t+1)}, \hat{\boldsymbol{\kappa}}^{(t+1)}$ ((2.22))

$$\begin{aligned} \min_{\mathbf{v},\hat{\boldsymbol{\kappa}}} \; & L(\boldsymbol{w}^{(t+1)}, \mathbf{K}^{(t+1)}, \boldsymbol{\kappa}^{(t+1)}, \mathbf{v}, \hat{\boldsymbol{\kappa}}, \boldsymbol{\mu}^{(t)}, \boldsymbol{\lambda}^{(t)}, \boldsymbol{\delta}^{(t)}) \\ \text{s.t. } & \hat{\kappa}_{u_k} = \sigma_{u_k}, \quad \forall u_k \in \mathcal{U}; \end{aligned}$$

5: Update the multipliers by

$$\begin{aligned} \mu_{u_k}^{j_l(t+1)} &= \mu_{u_k}^{j_l(t)} + \rho\left(K_{u_k}^{j_l(t+1)} - (\mathbf{h}_{u_k}^{l})^H \mathbf{v}_{j_l}^{(t+1)}\right), \; \forall \; u_k, j_l \in \mathcal{U} \\ \boldsymbol{\lambda}^{b_k(t+1)} &= \boldsymbol{\lambda}^{b_k(t)} + \rho(\boldsymbol{w}^{b_k(t+1)} - \mathbf{v}^{b_k(t+1)}), \quad \forall b_k \in \mathcal{B} \\ \delta_{u_k}^{(t+1)} &= \delta_{u_k}^{(t)} + \rho(\kappa_{u_k}^{(t+1)} - \hat{\kappa}_{u_k}^{(t+1)}), \quad \forall b_k \in \mathcal{B}; \end{aligned}$$

6: $t = t + 1$

7: **Until** Desired stopping criteria is met

Table 2.1: Summary of the proposed Algorithm 1

and $\mathbf{z} = [\mathbf{v}_{stack}^H, \hat{\boldsymbol{\kappa}}_{stack}^H]^H$ we can observe that

$$f(\boldsymbol{x}) = \sum_{b_k \in \mathcal{B}} \beta_{b_k} \|\boldsymbol{w}^{b_k}\|_2,\ g(\mathbf{z}) = \theta \sum_{b_k \in \mathcal{B}} \|\mathbf{v}^{b_k}\|_2^2,\ \mathbf{A} = \mathbf{I},\ \mathbf{B} = -\begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{H}_{stack} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix},\ \mathbf{c} = \mathbf{0}$$

$$\mathcal{C}_1 = \Bigg\{ \boldsymbol{x} \mid \|\boldsymbol{w}^{b_k}\|_2^2 \le P_{b_k},\ \ \forall\ b_k \in \mathcal{B}, \mathrm{Im}(K_{u_k}^{u_k}) = 0,$$

$$|K_{u_k}^{u_k}| \ge \sqrt{\tau_{u_k} \left(\kappa_{u_k}^2 + \sum_{(l,j) \ne (k,u)} |K_{u_k}^{j_l}|^2 \right)},\ \forall\ u_k \in \mathcal{U}, \Bigg\},$$

$$\mathcal{C}_2 = \{ \mathbf{z} \mid \hat{\kappa}_{u_k} = \sigma_{u_k},\ \ \forall u_k \in \mathcal{U} \},$$

where $\mathbf{H}_{stack} \in \mathbb{C}^{|\mathcal{U}|^2 \times N|\mathcal{B}||\mathcal{U}|}$ is a stacked matrix of $\{(\mathbf{h}_{j_l}^k)^H \mid j_l \in \mathcal{U}, k \in \mathcal{K}\}$ and $\mathbf{0}$'s in a way that $\mathbf{K}_{stack} - \mathbf{H}_{stack}\mathbf{v}_{stack} = \mathbf{0}$ is equivalent to (2.10a).

Since $\mathbf{A}^T\mathbf{A} = \mathbf{I}$ and $\mathbf{B}^T\mathbf{B} = \begin{bmatrix} \mathbf{I} + \mathbf{H}_{stack}^T\mathbf{H}_{stack} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}$ are invertible, and both $\mathcal{C}_1$ and $\mathcal{C}_2$ are convex sets, then by Proposition 1 in Appendix A, we can conclude that every limit point $\mathbf{v}^{(t)}$ (or $\boldsymbol{w}^{(t)}$) of Algorithm 1 is an optimal solution of problem (2.8). □

Since the formulated problem (2.8) is strongly convex. Algorithm 1 converges to the unique optimal solution of problem (2.8).

2.2.3 Step-by-Step Computation for the Proposed Algorithm

In the following, we will explain in detail how each primal variables $\boldsymbol{w}, \mathbf{K}, \boldsymbol{\kappa}, \mathbf{v}$, and $\hat{\boldsymbol{\kappa}}$ (ignoring the superscript iteration index for simplicity) is updated. As will be seen shortly, the update for the first block $\{\boldsymbol{w}, \mathbf{K}, \boldsymbol{\kappa}\}$ can be further decomposed into two independent problems, one for $\boldsymbol{w}$, and one for $\{\mathbf{K}, \boldsymbol{\kappa}\}$.

(1) Update $\{\mathbf{K}, \boldsymbol{\kappa}\}$: First observe that the subproblem related to $\{\mathbf{K}, \boldsymbol{\kappa}\}$ is independent of $\boldsymbol{w}$, and can be decoupled over each user. Therefore it can be written as $|\mathcal{U}|$

separate problems, with u_k-th subproblem expressed as

$$\begin{aligned}\min_{\{K_{u_k}^{j_l}\}_{j_l\in\mathcal{U},\kappa_{u_k}}} \quad & \mathrm{Re}\left(\sum_{j_l\in\mathcal{U}}\langle K_{u_k}^{j_l}-(\mathbf{h}_{u_k}^{l})^H\mathbf{v}_{j_l},\mu_{u_k}^{j_l}\rangle\right)+\delta_{u_k}(\kappa_{u_k}-\hat{\kappa}_{u_k})\\ & +\frac{\rho}{2}\sum_{j_l\in\mathcal{U}}\left|K_{u_k}^{j_l}-(\mathbf{h}_{u_k}^{l})^H\mathbf{v}_{j_l}\right|^2+\frac{\rho}{2}(\kappa_{u_k}-\hat{\kappa}_{u_k})^2\\ \text{s.t.} \quad & |K_{u_k}^{u_k}|\geq\sqrt{\tau_{u_k}\left(\kappa_{u_k}^2+\sum_{(l,j)\neq(k,u)}|K_{u_k}^{j_l}|^2\right)},\\ & \mathrm{Im}(K_{u_k}^{u_k})=0. \end{aligned}\tag{2.13}$$

By completing the squares, this problem can be equivalently written as

$$\begin{aligned}\min_{\{K_{u_k}^{j_l}\}_{j_l\in\mathcal{U},\kappa_{u_k}}} \quad & \left(\kappa_{u_k}-\hat{\kappa}_{u_k}+\frac{\delta_{u_k}}{\rho}\right)^2+\sum_{j_l\in\mathcal{U}}\left|K_{u_k}^{j_l}-(\mathbf{h}_{u_k}^{l})^H\mathbf{v}_{j_l}+\frac{\mu_{u_k}^{j_l}}{\rho}\right|^2\\ \text{s.t.} \quad & |K_{u_k}^{u_k}|\geq\sqrt{\tau_{u_k}\left(\kappa_{u_k}^2+\sum_{(l,j)\neq(k,u)}|K_{u_k}^{j_l}|^2\right)},\\ & \mathrm{Im}(K_{u_k}^{u_k})=0. \end{aligned}\tag{2.14}$$

Let us use $\{\{K_{u_k}^{j_l\star}\}_{j_l\in\mathcal{U}},\kappa_{u_k}^{\star}\}$ to denote the optimal solution of problem (2.14). Then the corresponding first-order optimality conditions are given by

$$K_{u_k}^{u_k\star}=\frac{1}{2}\gamma^{\star}+\mathrm{Re}\left((\mathbf{h}_{u_k}^{k})^H\mathbf{v}_{u_k}-\frac{\mu_{u_k}^{u_k}}{\rho}\right)\tag{2.15a}$$

$$K_{u_k}^{j_l\star}=\frac{\bar{K}_{u_k}\left((\mathbf{h}_{u_k}^{l})^H\mathbf{v}_{j_l}-\frac{\mu_{u_k}^{j_l}}{\rho}\right)}{\bar{K}_{u_k}+\frac{\gamma^{\star}\sqrt{\tau_{u_k}}}{2}},\quad \forall j_l\in\mathcal{U}, j_l\neq u_k\tag{2.15b}$$

$$\kappa_{u_k}^{\star}=\frac{\bar{K}_{u_k}\left(\hat{\kappa}_{u_k}-\frac{\delta_{u_k}}{\rho}\right)}{\bar{K}_{u_k}+\frac{\gamma^{\star}\sqrt{\tau_{u_k}}}{2}}\tag{2.15c}$$

$$K_{u_k}^{u_k\star}\geq\sqrt{\tau_{u_k}}\bar{K}_{u_k},\ \gamma^{\star}\geq 0,\ \left(K_{u_k}^{u_k\star}-\sqrt{\tau_{u_k}}\bar{K}_{u_k}\right)\gamma^{\star}=0\tag{2.15d}$$

where $\gamma^{\star}$ is the optimal Lagrangian dual variable for the second-order cone constraint of problem (2.14) and $\bar{K}_{u_k}\triangleq\sqrt{\kappa_{u_k}^{\star 2}+\sum_{(l,j)\neq(k,u)}\left|K_{u_k}^{j_l\star}\right|^2}$. If $\gamma^{\star}=0$, the objective value

of problem (2.14) is the minimum possible value, 0, and by complementarity condition (2.15d), this is possible only if

$$\left|\mathrm{Re}\left((\mathbf{h}_{u_k}^k)^H\mathbf{v}_{u_k}-\frac{\mu_{u_k}^{u_k}}{\rho}\right)\right|\geq\sqrt{\tau_{u_k}}\sqrt{(\hat{\kappa}_{u_k}-\frac{\delta_{u_k}}{\rho})^2+\sum_{(l,j)\neq(k,u)}\left|(\mathbf{h}_{u_k}^l)^H\mathbf{v}_{j_l}-\frac{\mu_{u_k}^{j_l}}{\rho}\right|^2}\triangleq\underline{K}_{u_k}. \tag{2.16}$$

On the other hand, if (2.16) does not hold, we know that $\gamma^\star\neq 0$, and by complementarity condition (2.15d), $\mathrm{Re}(K_{u_k}^{u_k\star})=\sqrt{\tau_{u_k}}\bar{K}_{u_k}$ holds. Therefore, the optimal dual variable, $\gamma^\star$ can be analytically solved as

$$\gamma^\star=2\frac{\underline{K}_{u_k}-\mathrm{Re}\left((\mathbf{h}_{u_k}^k)^H\mathbf{v}_{u_k}-\frac{\mu_{u_k}^{u_k}}{\rho}\right)}{1+\tau_{u_k}}$$

Hence, the optimal solution of problem (2.14) can be solved in closed-form by (2.15a), (2.15b), and (2.15c) with given $\gamma^\star$ and the fact that $\bar{K}_{u_k}=\mathrm{Re}(K_{u_k}^{u_k\star})/\sqrt{\tau_{u_k}}$.

It is worth noting that, this closed-form update rule is made possible by making κ_{u_k} as an optimization variable. This is the reason that we want to introduce extra variables $\{\kappa_{i_k}\}$ and $\{\hat{\kappa}_{i_k}\}$ in (2.10c).

(2) Update $\{\boldsymbol{w}\}$: The subproblem for the optimization variable $\boldsymbol{w}$ can also be decoupled over $|\mathcal{B}|$ separate subproblems, one for each BS q_k:

$$\begin{aligned}\min_{\boldsymbol{w}^{b_k}}\ & \beta_{b_k}\|\boldsymbol{w}^{b_k}\|_2+\frac{\rho}{2}\|\boldsymbol{w}^{b_k}-\mathbf{v}^{b_k}-\boldsymbol{\lambda}^{b_k}/\rho\|_2^2\\ \text{s.t.}\quad & \|\boldsymbol{w}^{b_k}\|_2^2\leq P_{b_k}.\end{aligned} \tag{2.17}$$

By defining $\mathbf{b}^{b_k}=\mathbf{v}^{b_k}+\boldsymbol{\lambda}^{b_k}/\rho$, the optimal solution $\boldsymbol{w}^{b_k\star}$ should satisfy the first-order optimality condition

$$\rho\mathbf{b}^{b_k}-\boldsymbol{w}^{b_k\star}(\rho+2\gamma^{b_k\star})\in\partial(\beta_{b_k}\|\boldsymbol{w}^{b_k\star}\|_2) \tag{2.18a}$$

$$\|\boldsymbol{w}^{b_k\star}\|_2^2\leq P_{b_k},\ \gamma^{b_k\star}\geq 0 \tag{2.18b}$$

$$(\|\boldsymbol{w}^{b_k\star}\|_2^2-P_{b_k})\gamma^{b_k\star}=0 \tag{2.18c}$$

where $\gamma^{b_k\star}$ is the optimal Lagrangian multiplier associated with the quadratic constraint $\|\boldsymbol{w}^{b_k}\|_2^2\leq P_{b_k}$. From (2.18a) and the definition of the subgradient for the ℓ_2 norm, we

have that $\boldsymbol{w}^{b_k\star} = \mathbf{0}$ whenever $\rho\|\mathbf{b}^{b_k}\|_2 \leq \beta_{b_k}$. When $\rho\|\mathbf{b}^{b_k}\|_2 > \beta_{b_k}$, we have

$$\begin{aligned} \rho\mathbf{b}^{b_k} - \boldsymbol{w}^{b_k\star}(\rho + 2\gamma^{b_k\star}) &= \beta_{b_k}\frac{\boldsymbol{w}^{b_k\star}}{\|\boldsymbol{w}^{b_k\star}\|_2} \\ \implies \boldsymbol{w}^{b_k\star} &= \frac{\mathbf{b}^{b_k}(\rho\|\mathbf{b}^{b_k}\|_2 - \beta_{b_k})}{(\rho + 2\gamma^{b_k\star})\|\mathbf{b}^{b_k}\|_2}. \end{aligned} \tag{2.19}$$

By the complementarity condition, $\gamma^{b_k\star} = 0$ if $\left\|\frac{\mathbf{b}^{b_k}(\rho\|\mathbf{b}^{b_k}\|_2 - \beta_{b_k})}{\rho\|\mathbf{b}^{b_k}\|_2}\right\|_2^2 \leq P_{b_k}$. Otherwise, $\gamma^{b_k\star}$ should be chosen such that $\|\boldsymbol{w}^{b_k\star}\|_2^2 = P_{b_k}$, which implies that $\gamma^{b_k\star} = (\rho(\|\mathbf{b}^{b_k}\|_2 - \sqrt{P_{b_k}}) - \beta_{b_k})/(2\sqrt{P_{b_k}})$. Plugging these choices of $\gamma^{b_k\star}$ into (2.19), then we conclude that the solution for problem (2.17) is given by

$$\boldsymbol{w}^{b_k\star} = \begin{cases} \mathbf{0}, & \rho\|\mathbf{b}^{b_k}\| \leq \beta_{b_k}, \\ \frac{\mathbf{b}^{b_k}(\rho\|\mathbf{b}^{b_k}\|_2 - \beta_{b_k})}{\rho\|\mathbf{b}^{b_k}\|_2}, & \rho\|\mathbf{b}^{b_k}\| > \beta_{b_k} \text{ and } \left\|\frac{\mathbf{b}^{b_k}(\rho\|\mathbf{b}^{b_k}\|_2 - \beta_{b_k})}{\rho\|\mathbf{b}^{b_k}\|_2}\right\|_2^2 \leq P_{b_k}, \\ \sqrt{P_{b_k}}\frac{\mathbf{b}^{b_k}}{\|\mathbf{b}^{b_k}\|_2}, & \text{otherwise.} \end{cases} \tag{2.20}$$

(3) Update $\mathbf{v}, \hat{\boldsymbol{\kappa}}$: From step 4 of Algorithm 1, we readily have $\hat{\kappa}_{u_k}^\star = \sigma_{u_k}, \ \forall u_k \in \mathcal{U}$. The subproblem for the block variable $\mathbf{v}$ can be written as K independent unconstrained quadratic problems, one for each cell k:

$$\begin{aligned} \min_{\{\mathbf{v}^{b_k}\}_{b_k\in\mathcal{B}_k}} \frac{\rho}{2}\sum_{\substack{u_k\in\mathcal{U}_k \\ j_l\in\mathcal{U}}}\left|(\mathbf{h}_{j_l}^k)^H\mathbf{v}_{u_k} - K_{j_l}^{u_k} - \mu_{j_l}^{u_k}/\rho\right|^2 + \theta\sum_{b_k\in\mathcal{B}_k}\|\mathbf{v}^{b_k}\|_2^2 \\ + \frac{\rho}{2}\sum_{b_k\in\mathcal{B}_k}\|\mathbf{v}^{b_k} - \boldsymbol{w}^{b_k} + \boldsymbol{\lambda}^{b_k}/\rho\|_2^2. \end{aligned} \tag{2.21}$$

The solution for this unconstrained problem is given by

$$\mathbf{v}_{u_k}^\star = \rho^{-1}\left((1 + 2\theta/\rho)\mathbf{I} + \mathbf{H}^k\mathbf{H}^{kH}\right)^{-1}(\rho\mathbf{H}^k\mathbf{K}^{u_k} + \mathbf{H}\boldsymbol{\mu}^{u_k} + \rho\boldsymbol{w}_{u_k} - \boldsymbol{\lambda}_{u_k}), \ \forall u_k \in \mathcal{U}_k \tag{2.22}$$

where $\mathbf{H}^k = \left[\{\mathbf{h}_{j_l}^k\}_{j_l\in\mathcal{U}}\right] \in \mathbb{C}^{NB_k\times|\mathcal{U}|}$, $\mathbf{K}^{u_k} = \left[\{\mathbf{K}_{j_l}^{u_k}\}_{j_l\in\mathcal{U}}\right]^T \in \mathbb{C}^{|\mathcal{U}|}$, $\boldsymbol{\mu}^{u_k} = [\{\mu_{j_l}^{u_k}\}_{j_l\in\mathcal{U}}]^T \in \mathbb{C}^{|\mathcal{U}|}$, and $\boldsymbol{\lambda}_{u_k} = [(\boldsymbol{\lambda}_{u_k}^{1_k})^T, \ldots, (\boldsymbol{\lambda}_{u_k}^{B_k})^T]^T \in \mathbb{C}^{NB_k}$, with $\boldsymbol{\lambda}_{u_k}^{b_k} \in \mathbb{C}^N$ being the u_k-th block of $\boldsymbol{\lambda}^{b_k}$. Hence, the optimization variable block $\mathbf{v}$ can be optimally solved in closed-form as well.

2.2.4 Discussions

Computational Costs

As noted above, each step of Algorithm1 can be carried out in closed-form, which makes Algorithm 1 highly efficient. Specifically, the most computational intensive operation in Algorithm 1 is the matrix inversion (2.22), which has complexity in the order of $\mathrm{O}((NB_k)^3)$. However, this operation only needs to be computed once for each cell k. As a result, compared to the standard interior point algorithm, which has a per iteration complexity in the order $\mathrm{O}((\sum_{k\in\mathcal{K}} NB_kU_k)^3)$, the proposed ADMM approach has a cheaper per iteration computational cost, especially when $|\mathcal{B}|$ and $|\mathcal{U}|$ are large.

Distributed Implementation

Another advantage of the proposed algorithm is that it can be implemented without a central controller. Observe that except for $\{\mathbf{K}, \boldsymbol{\kappa}\}$, the computation for the rest of the primal and dual variables can be performed within each cell without any information exchange among the cells. When updating $\mathbf{K}$ and $\boldsymbol{\kappa}$, each cell k exchanges $|\mathcal{U}_k||\mathcal{U}|$ complex values $\{(\mathbf{h}_{j_l}^k)^H \mathbf{v}_{u_k} | j_l \in \mathcal{U}, u_k \in \mathcal{U}_k\}$ with the rest of cells. This is possible since each cell operator k can collect the locally estimated channel information $\{\mathbf{h}_{j_l}^k | \mid j_l \in \mathcal{U}\}$ from BSs via backhaul links or control channels. Once this is done, the subproblems (2.14) for updating $\{\mathbf{K}, \boldsymbol{\kappa}\}$ can be again solved independently and simultaneously by each cell operator. In conclusion, the ADMM approach allows problem (2.8) to be solved in a distributed manner across cells without a central operator.

The Debiasing Step

After problem (2.8) is solved, performing an additional "de-biasing" step can further minimize the total power consumption. That is, with the given set of selected active BSs computed by the proposed single-stage ADMM approach, we can solve problem (2.8) again, this time *without* the sparse promoting terms. This can be done by making the following changes to the proposed algorithm: *1)* letting $\beta_{b_k} = 0, \ \forall b_k \in \mathcal{B}$; *2)* setting $\theta = 1$; *3)* only optimize over BSs with $\mathbf{v}^{b_k\star} \neq 0$. See reference [114] for further justification of using such de-biasing technique in solving regularized optimization problems.

The Special Case of Power Minimization Problem

As a byproduct of the proposed ADMM approach, the conventional power minimization problem (2.4) without active BS selection can also be efficiently solved using a simplified version of Algorithm 1, by setting $\beta_{b_k} = 0, \quad \forall b_k \in \mathcal{B}$, and $\theta = 1$. Compared to the existing approaches for solving the same problem, the proposed ADMM approach is computationally more efficient. For example, the uplink-downlink duality approach [29] needs to perform matrix inversion operations with complexity $\mathrm{O}((NB_k)^3)$ in each iteration. The other ADMM based algorithms for solving problem (2.4) either needs to solve SDPs [36] or SOCPs [96] in each iteration.

In contrast, by a novel splitting of the primal variables according to the special structure of (2.4), our proposed ADMM approach (i.e., Algorithm 1) does not solve expensive subproblems; the subproblems are all solvable in closed forms.

As a remark, if there are multiple-antennas at both BSs and users, the corresponding MIMO beamforming design for power minimization becomes nonconvex. In this case, the ADMM is not applicable. However, as in many existing works, e.g., [115], one can update the transmit and receive beamformers alternately, resulting in convex subproblems that can still be solved by the ADMM.

Further Reduction of the Number of Active BSs

To achieve the maximum reduction of the number of active BSs, we propose to adaptively reweight the coefficients $\beta_{b_k}, \quad \forall b_k \in \mathcal{B}$. This reweighting technique is popular in the compressive sensing literature to increase the sparsity level of the solution; see e.g., [41, 105]. This can be done by first solving problem (2.8), and then updating the coefficient β_{b_k} by

$$\beta_{b_k} \longleftarrow \frac{\beta_{b_k}^{(0)}}{\|\boldsymbol{w}^{b_k\star}\| + \epsilon}, \quad \forall b_k \in \mathcal{B}, \tag{2.23}$$

where $\beta_{b_k}^{(0)}, \forall b_k \in \mathcal{B}$, are the initial β_{b_k} of problem (2.8) and $\epsilon > 0$ is a small prescribed parameter to provide the stability when $\|\boldsymbol{w}^{b_k\star}\|$ is too small. With this new set of β_{b_k}, (2.8) is solved again. Intuitively, those BFs that have smaller magnitude will be penalized more heavily in the comming iteration, thus is more likely to be set to zero. In our numerical experiments to be shown in Sec. 2.4, indeed we observe that by using

such reweighting technique, the number of active BSs converges very fast and is much smaller than that obtained by solving problem (2.8) only once.

2.3 Sum Rate Maximization with Base Station Activation

2.3.1 Problem Formulation

In this section, we show that how BS activation can be incorporated into the design criteria **C2)**, i.e., maximize the sum rate subject to power constraint. We first note that, as explained in Sec. 2.1, even without considering BS activation, solving sum rate maximization problem (2.5) is itself strongly NP-hard. Since this problem remains NP-hard regardless the number of antennas at each user, we will consider a more general scenario in which both BSs and users are equipped with multiple antennas.

For simplicity of notation, we assume that all users have L receive antennas. Let us change the notation of channel from $\mathbf{h}_{u_k}^{ql}$ to $\mathbf{H}_{u_k}^{ql} \in \mathbb{C}^{L\times N}$. In this way, the achievable rate for user u_k becomes

$$R_{u_k}(\mathbf{v}) = \log\det\left(\mathbf{I} + \mathbf{H}_{u_k}^{k}\mathbf{v}_{u_k}\mathbf{v}_{u_k}^{H}(\mathbf{H}_{u_k}^{k})^{H}\big(\sum_{(l,j)\neq(k,u)}\mathbf{H}_{u_k}^{l}\mathbf{v}_{jl}\mathbf{v}_{jl}^{H}(\mathbf{H}_{u_k}^{l})^{H} + \sigma_{u_k}^{2}\mathbf{I}\big)^{-1}\right). \tag{2.24}$$

Similar to the previous section, we aim at jointly maximizing the sum rate and selecting the active BSs. To this end, we first split the transmit BF $\mathbf{v}_{u_k}^{b_k}$ by $\mathbf{v}_{u_k}^{b_k} = \alpha_{b_k}\bar{\mathbf{v}}_{u_k}^{b_k}$, with $\alpha_{b_k} \in [0,1]$ representing whether BS b_k is switched on. That is, when $\alpha_{b_k} = 0$, BS b_k is switched off, otherwise, BS b_k is turned on. In the sequel, we will consider the following single-stage regularized sum rate maximization problem

$$\begin{aligned}\max_{\boldsymbol{\alpha},\bar{\mathbf{v}}}\ & \sum_{k\in\mathcal{K}}\sum_{u_k\in\mathcal{U}_k} R_{u_k}(\mathbf{v}) - \sum_{b_k\in\mathcal{B}}\mu_{b_k}\|\alpha_{b_k}\|_0 \\ \text{s.t. }& \alpha_{b_k}^2(\bar{\mathbf{v}}^{b_k})^{H}\bar{\mathbf{v}}^{b_k} \le P_{b_k},\quad \forall b_k\in\mathcal{B}_k,\ k\in\mathcal{K},\end{aligned} \tag{2.25}$$

where $\mu_{b_k} \ge 0, \quad \forall b_k \in \mathcal{B}$, is the parameter controlling the number of active BSs; $\boldsymbol{\alpha} \triangleq \{\boldsymbol{\alpha}_k \mid k\in\mathcal{K}\}$ with $\boldsymbol{\alpha}_k \triangleq [\alpha_{1_k}, \alpha_{2_k}, \ldots, \alpha_{B_k}]^T \in \mathbb{R}^{B_k}$.

Unfortunately, the ℓ_0 norm is not only non-convex but also not continuous. As a result it is difficult to find even a locally optimal solution for problem (2.25). Similar

to the previous section, we will relax, in the following, the ℓ_0 norm to the ℓ_1 norm. In the way, the regularized sum rate maximization problem becomes

$$\begin{aligned}\max_{\boldsymbol{\alpha},\bar{\mathbf{v}}} \; & \sum_{k\in\mathcal{K}}\sum_{u_k\in\mathcal{U}_k} R_{u_k}(\mathbf{v}) - \sum_{b_k\in\mathcal{B}} \mu_{b_k}|\alpha_{b_k}| \\ \text{s.t. } & \alpha_{b_k}^2(\bar{\mathbf{v}}^{b_k})^H\bar{\mathbf{v}}^{b_k} \le P_{b_k}, \quad \forall b_k\in\mathcal{B}_k,\ k\in\mathcal{K}, \end{aligned} \tag{2.26}$$

In what follows, we will propose an efficient algorithm to compute a stationary solution for this relaxed problem.

Remark 1 *Instead of splitting* $\mathbf{v}_{u_k}^{b_k}$ *and penalizing* $\|\boldsymbol{\alpha}_k\|_1$*, another natural modification is to add a group LASSO (Least Absolute Shrinkage and Selection Operator) regularization term (e.g., [40]) for each BS's BF directly, i.e., use the regularization term* $\|\mathbf{v}^{b_k}\|$ *for BS* b_k *in the objective function of problem* (2.5)*. However, when the power used by BS* b_k *is large, the magnitude of penalization term can dominate that of the system sum rate. Thus solving such group-LASSO penalized problem would effectively force the BSs to use only a small portion of its power budget, which could lead to a dramatic reduction of the system sum rate. The regularization in* (2.26) *avoids this problem.*

Remark 2 *For simplicity and consistency, we have focused on the vector beamformer case (i.e., one data stream per receiver) in this section. It should be noted that all the results derived in this section can be straightforwardly extended to the matrix precoder case [116].*

2.3.2 Active BS Selection via a Sparse WMMSE Algorithm

By using a similar argument as in [21, Proposition 1], we can show that the penalized sum rate maximization problem (2.26) is equivalent to the following penalized weighted mean square error (MSE) minimization problem

$$\min_{\boldsymbol{\alpha},\bar{\mathbf{v}},\mathbf{u},\mathbf{w}} f(\mathbf{v},\mathbf{w},\mathbf{u}) + \sum_{b_k\in\mathcal{B}} \mu_{b_k}|\alpha_{b_k}| \tag{2.27a}$$

$$\text{s.t. } f(\mathbf{v},\mathbf{w},\mathbf{u}) = \sum_{u_k\in\mathcal{U}} w_{u_k} e_{u_k}(\mathbf{u}_{u_k},\mathbf{v}) - \log(w_{u_k}) \tag{2.27b}$$

$$\alpha_{b_k}^2(\bar{\mathbf{v}}^{b_k})^H\bar{\mathbf{v}}^{b_k} \le P_{b_k}, \quad \forall b_k\in\mathcal{B}_k,\ k\in\mathcal{K}. \tag{2.27c}$$

In the above expression, $\mathbf{u} \triangleq \{\mathbf{u}_{u_k} \mid u_k \in \mathcal{U}\}$ is the set of all receive BFs of the users; $\mathbf{w} \triangleq \{w_{u_k} | u_k \in \mathcal{U}\}$ is the set of non-negative weights; e_{u_k} is the MSE for estimating s_{u_k}:

$$e_{u_k}(\mathbf{u}_{u_k},\mathbf{v}) \triangleq (1 - \mathbf{u}_{u_k}^H \mathbf{H}_{u_k}^k \mathbf{v}_{u_k})(1 - \mathbf{u}_{u_k}^H \mathbf{H}_{u_k}^k \mathbf{v}_{u_k})^H + \sum_{(\ell,j)\neq(k,u)} \mathbf{u}_{u_k}^H \mathbf{H}_{u_k}^\ell \mathbf{v}_{j_\ell} \mathbf{v}_{j_\ell}^H (\mathbf{H}_{u_k}^\ell)^H \mathbf{u}_{u_k} + \sigma_{u_k}^2 \mathbf{u}_{u_k}^H \mathbf{u}_{u_k}.$$

To guarantee convergence of the proposed algorithm, we further replace the power constraint (2.27c) by a slightly more conservative constraint, namely $(\bar{\mathbf{v}}^{b_k})^H \bar{\mathbf{v}}^{b_k} \leq P_{b_k},\ \alpha_{b_k}^2 \leq 1$. The precise reason for doing so will be explained shortly in the reasoning of Theorem 3. In this way, the modified penalized weighted MSE minimization problem for active BS selection is given by

$$\begin{aligned} \min_{\boldsymbol{\alpha},\bar{\mathbf{v}},\mathbf{u},\mathbf{w}} & f(\mathbf{v},\mathbf{w},\mathbf{u}) + \sum_{b_k \in \mathcal{B}} \mu_{b_k} |\alpha_{b_k}| \\ \text{s.t. } & f(\mathbf{v},\mathbf{w},\mathbf{u}) = \sum_{u_k \in \mathcal{U}} w_{u_k} e_{u_k}(\mathbf{u}_{u_k},\mathbf{v}) - \log(w_{u_k}) \\ & (\bar{\mathbf{v}}^{b_k})^H \bar{\mathbf{v}}^{b_k} \leq P_{b_k},\ \alpha_{b_k}^2 \leq 1, \quad \forall b_k \in \mathcal{B}_k,\ k \in \mathcal{K}. \end{aligned} \tag{2.28}$$

Although the modified power constraint will shrink the original feasible set whenever $\alpha_{b_k}^2 \neq 0$ or ± 1, thus may reduce the sum rate performance of the obtained transceiver, our numerical experiments (to be shown in Section 2.4) suggest that satisfactory sum rate performance can still be achieved.

Because of the nonconvexity, it is difficult to solve (2.28) to global optimality. In the following, we propose an efficient algorithm that can at least solve the problem to a stationary solution. Due to the fact that problem (2.28) is convex in each block variables, global minimum can be obtained for each block variable when fixing the rest. Furthermore, the problem is strongly convex for block $\mathbf{u}$ and $\mathbf{w}$, respectively, and the unique optimal solution $\mathbf{u}_{u_k}^\star$ and $w_{u_k}^\star,\ \forall u_k \in \mathcal{U}$, can be obtained in closed-form:

$$\mathbf{u}_{u_k}^\star(\mathbf{v}) = \left(\sum_{j_l \in \mathcal{U}} \mathbf{H}_{u_k}^l \mathbf{v}_{j_l} \mathbf{v}_{j_l}^H (\mathbf{H}_{u_k}^l)^H + \sigma_{u_k}^2 \mathbf{I} \right)^{-1} \mathbf{H}_{u_k}^k \mathbf{v}_{u_k} \triangleq \mathbf{J}_{u_k}^{-1}(\mathbf{v}) \mathbf{H}_{u_k}^k \mathbf{v}_{u_k} \tag{2.29}$$

$$w_{u_k}^\star(\mathbf{v}) = \left(1 - \mathbf{v}_{u_k}^H \left(\mathbf{H}_{u_k}^k \right)^H \mathbf{J}_{u_k}^{-1}(\mathbf{v}) \mathbf{H}_{u_k}^k \mathbf{v}_{u_k} \right)^{-1}. \tag{2.30}$$

On the other hand, problem (2.28) can also be rewritten as

$$\min_{\boldsymbol{\alpha},\bar{\mathbf{v}},\mathbf{u},\mathbf{w}} f(\mathbf{v},\mathbf{w},\mathbf{u}) + \sum_{b_k\in\mathcal{B}} \mu_{b_k}|\alpha_{b_k}| + I_1(\bar{\mathbf{v}}) + I_2(\boldsymbol{\alpha}) \tag{2.31}$$

where $I_1(\bar{\mathbf{v}})$ and $I_2(\boldsymbol{\alpha})$ are indicator functions for both constraints defined respectively as

$$I_1(\bar{\mathbf{v}}) = \begin{cases} 0, & \text{if } (\bar{\mathbf{v}}^{b_k})^H \bar{\mathbf{v}}^{b_k} \le P_{b_k},\ \forall b_k \in \mathcal{B}_k, k \in \mathcal{K}, \\ \infty, & \text{otherwise} \end{cases},$$

$$I_2(\boldsymbol{\alpha}) = \begin{cases} 0, & \text{if } \alpha_{b_k}^2 \le 1,\ \ \forall b_k \in \mathcal{B}_k,\ k \in \mathcal{K}, \\ \infty, & \text{otherwise} \end{cases}.$$

Observe that when the problem is written in the form of (2.31), all its nonsmooth parts are *separable* across block variables $\boldsymbol{\alpha}$, $\bar{\mathbf{v}}$, $\mathbf{u}$, and $\mathbf{w}$. Such separability is guaranteed by our modified power constraints, and is referred to as the "regularity condition" for nonsmooth optimization; see [117] for details about this condition. Combining this property with the fact that at most two blocks, namely $\boldsymbol{\alpha}$ and $\bar{\mathbf{v}}$, may not have unique minimizer, a block coordinate descent (BCD) procedure [1] is guaranteed to converge to the stationary point of problem (2.28). This is proven by Lemma 3.1 and Theorem 4.1 of [117]. The following theorem summarizes the preceding discussion.

Theorem 3 *A BCD procedure that iteratively optimizes problem* (2.28) *for each block variables* $\mathbf{u}$, $\mathbf{w}$, $\bar{\mathbf{v}}$, *and* $\boldsymbol{\alpha}$, *can always converge to a stationary solution of problem* (2.28).

In the following, we discuss in detail how problem (2.28) can be solved for each block variables in an efficient manner. For blocks $\mathbf{u}$ and $\mathbf{w}$, optimal solutions are shown in (2.29) and (2.30), respectively. For the optimization problem of $\boldsymbol{\alpha}$, notice that when fixing $(\mathbf{u},\mathbf{w},\bar{\mathbf{v}})$, the objective of problem (2.28) is separable among the cells. Therefore K independent subproblems can be solved simultaneously, with the k-th subproblem assuming the following form

$$\begin{aligned} \min_{\boldsymbol{\alpha}_k}\ & (\boldsymbol{\alpha}_k)^T \mathbf{A}_k \boldsymbol{\alpha}_k - 2Re(\mathbf{b}_k^H \boldsymbol{\alpha}_k) + \sum_{b_k\in\mathcal{B}} \mu_{b_k}|\alpha_{b_k}| \\ \text{s.t. } & \alpha_{b_k}^2 \le 1, \quad \forall b_k \in \mathcal{B}_k \end{aligned} \tag{2.32}$$

[1] In our context, the BCD procedure refers to the computation strategy that cyclically updates the blocks $\mathbf{u}$, $\mathbf{w}$, $\bar{\mathbf{v}}$, and $\boldsymbol{\alpha}$ one at a time.

where

$$\mathbf{A}_k \triangleq \sum_{u_k \in \mathcal{U}_k} \mathrm{diag}(\bar{\mathbf{v}}_{u_k})^H \left(\sum_{j_l \in \mathcal{U}} w_{j_l} (\mathbf{H}_{j_l}^k)^H \mathbf{u}_{j_l} \mathbf{u}_{j_l}^H \mathbf{H}_{j_l}^k \right) \mathrm{diag}(\bar{\mathbf{v}}_{u_k})$$
$$\mathbf{b}_k \triangleq \sum_{u_k \in \mathcal{U}_k} w_{u_k} \mathrm{diag}(\bar{\mathbf{v}}_{u_k})^H (\mathbf{H}_{u_k}^k)^H \mathbf{u}_{u_k}.$$

Problem (2.32) is a quadratically constrained LASSO problem. It can be solved optimally by again applying a BCD procedure, with the block variables given by α_{b_k}, $\forall b_k \in \mathcal{B}_k$ (e.g., [114]). For the b_k-th block, its optimal solution $\alpha_{b_k}^\star$ must satisfy the following first-order optimality condition

$$2(c_{b_k} - (\mathbf{A}_k[b_k, b_k] + \gamma_{b_k}^\star)\alpha_{b_k}^\star) \in \mu_{b_k} \partial |\alpha_{b_k}^\star|, \tag{2.33a}$$
$$\gamma_{b_k}^\star \geq 0, \ (1 - (\alpha_{b_k}^\star)^2) \geq 0 \tag{2.33b}$$
$$(1 - (\alpha_{b_k}^\star)^2)\gamma_{b_k}^\star = 0 \tag{2.33c}$$

where $\gamma_{b_k}^\star$ is the optimal dual variable for the b_kth power constraint of problem (2.32), and $c_{b_k} \triangleq \mathrm{Re}(\mathbf{b}_k[b_k]) - \sum_{j_k \neq b_k} \mathbf{A}_k[j_k, b_k]\alpha_{j_k}$. Therefore, when $2\,|c_{b_k}| \leq \mu_{b_k}$, we have $\alpha_{b_k}^\star = 0$. In the following, let us focus on the case where $2|c_{b_k}| > \mu_{b_k}$. In this case, from the expression of the subgradient (2.33a), we have $\alpha_{b_k}^\star = \frac{-\mu_{b_k}\mathrm{sign}(\alpha_{b_k}^\star) + 2c_{b_k}}{2(\mathbf{A}_k[b_k,b_k] + \gamma_{b_k}^\star)}$. Since $\gamma_{b_k}^\star \geq 0$, $\mathbf{A}_k[b_k, b_k] \geq 0$, and $2|c_{b_k}| > \mu_{b_k}$, we have $\mathrm{sign}(\alpha_{b_k}^\star) = \mathrm{sign}(c_{b_k})$. By plugging $\alpha_{b_k}^\star$ into the objective function of problem (2.32), it can be shown the objective value is an increasing function of $\gamma_{b_k}^\star$. Therefore, by the monotonicity of $\gamma_{b_k}^\star$, primal and dual constraints (2.33b), and the complementarity condition (2.33c), in the case of $2|c_{b_k}| > \mu_{b_k}$, $\alpha_{b_k}^\star$ has the following structure

$$\alpha_{b_k}^\star = \begin{cases} \frac{-\mu_{b_k}\mathrm{sign}(c_{b_k}) + 2c_{b_k}}{2\mathbf{A}_k[b_k,b_k]}, & \text{if } \left| \frac{-\mu_{b_k}\mathrm{sign}(c_{b_k}) + 2c_{b_k}}{2\mathbf{A}_k[b_k,b_k]} \right| < 1 \\ \mathrm{sign}(c_{b_k}), & \text{otherwise} \end{cases} \tag{2.34}$$

Similarly, when fixing $(\boldsymbol{\alpha}, \mathbf{w}, \mathbf{u})$, the optimization problem for $\mathbf{v}$ is convex and separable among K cells, and the k-th subproblem is expressed as

$$\begin{aligned} \min_{\bar{\mathbf{v}}_{u_k},\ u_k \in \mathcal{U}_k} \quad & \sum_{u_k \in \mathcal{U}_k} \left(\bar{\mathbf{v}}_{u_k}^H \mathbf{C}_k \bar{\mathbf{v}}_{u_k} - \bar{\mathbf{v}}_{u_k}^H \mathbf{D}_{u_k} - \mathbf{D}_{u_k}^H \bar{\mathbf{v}}_{u_k} \right) \\ \text{s.t.} \quad & \sum_{u_k \in \mathcal{U}_k} (\bar{\mathbf{v}}_{u_k}^{b_k})^H \bar{\mathbf{v}}_{u_k}^{b_k} \leq P_{b_k}, \quad \forall b_k \in \mathcal{B}_k, \end{aligned} \tag{2.35}$$

where

$$\mathbf{C}_k \triangleq \hat{\boldsymbol{\alpha}}_k \left(\sum_{j_l \in \mathcal{U}} w_{j_l} (\mathbf{H}_{j_l}^k)^H \mathbf{u}_{j_l} \mathbf{u}_{j_l}^H \mathbf{H}_{j_l}^k \right) \hat{\boldsymbol{\alpha}}_k \in \mathbb{C}^{B_k N \times B_k N},$$
$$\mathbf{D}_{u_k} \triangleq w_{u_k} \hat{\boldsymbol{\alpha}}_k (\mathbf{H}_{u_k}^k)^H \mathbf{u}_{u_k} \in \mathbb{C}^{B_k N}, \quad \forall u_k \in \mathcal{U}_k,$$
$$\hat{\boldsymbol{\alpha}}_k \triangleq \mathrm{diag}(\alpha_{1_k} \mathbf{I}, \ldots, \alpha_{B_k} \mathbf{I}) \in \mathbb{C}^{B_k N \times B_k N}.$$

We wish to efficiently solve the problem by iteratively updating its block components $\bar{\mathbf{v}}^{b_k}, \ \forall b_k \in \mathcal{B}_k$. However, as discussed in Theorem 3, the algorithm convergence requires that the optimization problem has at most two block components which do not have unique optimal solution. To furfill this requirement, we add a regularization term $\sum_{b_k \in \mathcal{B}_k} \epsilon (\bar{\mathbf{v}}^{b_k})^H \bar{\mathbf{v}}^{b_k}$ to the objection function of problem (2.35) with $\epsilon > 0$. Thus, when $\epsilon \to 0$, the solution for the BF $\bar{\mathbf{v}}^{b_k \star}$ can be obtained by checking the first order optimality condition, and this can be expressed as

$$\bar{\mathbf{v}}_{u_k}^{b_k \star}(\delta_{b_k}) = \left(\mathbf{C}_k[b_k, b_k] + \delta_{b_k}^\star \mathbf{I} \right)^\dagger \left(\mathbf{D}_{u_k}[b_k] - \sum_{j_k \neq b_k} \mathbf{C}_k[b_k, j_k] \bar{\mathbf{v}}_{u_k}^{j_k \star} \right), \quad \forall u_k \in \mathcal{U}_k. \tag{2.36}$$

In the above expression, $\dagger$ denotes the Moore-Penrose pseudoinverse; $\delta_{b_k}^\star \geq 0$ is the optimal dual variable for the b_k-th power constraint; $\mathbf{C}_k[b_k, j_k] \in \mathbb{C}^{N \times N}$ and $\mathbf{D}_{u_k}[b_k] \in \mathbb{C}^N$ are, respectively, subblocks of matrices $\mathbf{C}_k$ and $\mathbf{D}_{u_k}$. By the complementarity condition, $\delta_{b_k}^\star = 0$ if $(\bar{\mathbf{v}}^{b_k \star}(0))^H \bar{\mathbf{v}}^{b_k \star}(0) \leq P_{b_k}$. Otherwise, it should satisfy $(\bar{\mathbf{v}}^{b_k \star}(\delta_{b_k}^\star))^H \bar{\mathbf{v}}^{b_k \star}(\delta_{b_k}^\star) = P_{b_k}$. For the latter case, $\delta_{b_k}^\star$ can be found by a simple bisection method.

In summary, our main algorithm can be summarized as S-WMMSE algorithm.

Sparse WMMSE (S-WMMSE) algorithm:

1: **Initialization** Generate a feasible set of variables $\{\bar{\mathbf{v}}_{u_k}\},\ u_k \in \mathcal{U}$, and let $\alpha_{b_k} = 1\ \ \forall b_k \in \mathcal{B}_k, k \in \mathcal{K}$.

2: **Repeat**

3: $\mathbf{u}_{u_k} \leftarrow \mathbf{J}_{u_k}^{-1}(\mathbf{v})\mathbf{H}_{u_k}^k \mathbf{v}_{u_k},\ \ \forall u_k \in \mathcal{U}$

4: $w_{u_k} \leftarrow (1 - \mathbf{v}_{u_k}^H \left(\mathbf{H}_{u_k}^k\right)^H \mathbf{J}_{u_k}^{-1}(\mathbf{v})\mathbf{H}_{u_k}^k \mathbf{v}_{u_k})^{-1},\ \ \forall u_k \in \mathcal{U}$

5: $\bar{\mathbf{v}}^{b_k}$ is iteratively updated by (2.36), $\forall b_k \in \mathcal{B}_k,\ \forall k \in \mathcal{K}$

6: α_{b_k} is iteratively updated by

$$\alpha_{b_k} = \begin{cases} 0, & \text{if } 2|c_{b_k}| \le \mu_{b_k} \\ (2.34), & \text{otherwise} \end{cases}, \quad \forall b_k \in \mathcal{B}_k,\ k \in \mathcal{K}$$

7: **Until** Desired stopping criteria is met

Similar to what we have done in the previous section, the de-biasing and reweighting procedures can further improve the sum rate performance and decrease the number of active BSs, respectively. The de-biasing procedure utilizes the given set of active BSs computed by the S-WMMSE algorithm, and solve problem (2.28) again, this time *without* the sparse promoting terms. In particular we make the following changes to the S-WMMSE algorithm: *1)* letting $\mu_{b_k} = 0$ for each $b_k \in \mathcal{B}$; *2)* skipping step 6; *3)* setting $\alpha_{b_k} = \text{sign}(\alpha_{b_k}^\star),\ \forall b_k$. In the reweighting procedure, we iteratively apply S-WMMSE to the reweighted problem with the parameter μ_{b_k} being updated by

$$\mu_{b_k} \longleftarrow \frac{\mu_{b_k}^{(0)}}{|\alpha_{b_k}| + \epsilon}, \quad \forall b_k \in \mathcal{B}, \tag{2.37}$$

where $\mu_{b_k}^{(0)}, \forall b_k \in \mathcal{B}$, are the initial μ_{b_k} of problem (2.28).

Furthermore, the proposed S-WMMSE algorithm can be solved distributively among each cell, under the following assumptions: i) there is a central controller in each cell; ii) the central controller for cell k has the CSI $\mathbf{H}_{j_l}^k,\ \forall j_l \in \mathcal{U}$, collected from BSs $b_k \in \mathcal{B}_k$ via high-speed backhaul links between BSs and iii) each user $u_k \in \mathcal{U}$ can locally estimate the received signal plus noise covariance matrix $\mathbf{J}_{u_k}$ and the received channel matrix $\mathbf{H}_{u_k}^k$. The last assumption ensures that user u_k can update $\mathbf{u}_{u_k}$ and w_{u_k} locally. After updating $\mathbf{u}_{u_k}$ and w_{u_k}, each user u_k can broadcast them to all the central controllers. Combined with assumption ii), the central controller in cell k can then update $\bar{v}^{b_k}$ and

$\alpha_{b_k}, \ \forall b_k \in \mathcal{B}_k$.

2.3.3 Joint active BS selection and BS clustering

In addition to controlling the number of active BSs, we can further optimize the size of BS clusters by adding an additional penalization on the BFs. Specifically, user u_k is not served by BS b_k when $\mathbf{v}_{u_k}^{b_k}$ is zero. It follows that user u_k is served with a small BS cluster if $\|\mathbf{v}_{u_k}^{b_k}\|$ is nonzero for only a few b_k's. Thus, a set of group LASSO regularization terms, $\sum_{b_k \in \mathcal{B}_k} \left\|\mathbf{v}_{u_k}^{b_k}\right\|$, $u_k \in \mathcal{U}$, can be added to the objective function of problem (2.5) to reduce the size of BS clusters; see [21] for details. Hence, to jointly control the size of BS cluster and reducing the BS usage, the objective function of the penalized weighted MMSE minimization problem (2.28) is now modified as

$$f(\mathbf{v}, \mathbf{w}, \mathbf{u}) + \sum_{k \in \mathcal{K}} \left(\sum_{u_k \in \mathcal{U}_k} \lambda_k \sum_{b_k \in \mathcal{B}_k} \|\bar{\mathbf{v}}_{u_k}^{b_k}\| \right) + \sum_{b_k \in \mathcal{B}} \mu_{b_k} |\alpha_{b_k}|, \tag{2.38}$$

where $\lambda_k \geq 0, \ \forall k \in \mathcal{K}$, is the parameter to control the size of BS cluster in cell k. For this modified problem, again a BCD procedure with block variables, $\boldsymbol{\alpha}$, $\bar{\mathbf{v}}$, $\mathbf{u}$, and $\mathbf{w}$, can be used to compute a locally optimal solution. The only difference from the algorithm proposed in the previous section is the computation of $\bar{\mathbf{v}}$. This can be carried out by solving a quadratically constrained group LASSO problem. See in [21, Table I] for details.

2.4 Numerical Experiments

In the following numerical experiments, we consider HetNets with at most 10 cells. The distance between the centers of adjacent cells is set as 2000 meters. In each cell $k \in \mathcal{K}$, we place one BS at the center of the cell (representing the macro BS), and place U users and $B-1$ remaining micro BSs randomly and uniformly; see Fig. 2.2 for an illustration of the network configuration. The channel model we use is Rayleigh channel with zero mean and variance $(200/d_{u_k}^{q_l})^3 L_{u_k}^{q_l}$, where $d_{u_k}^{q_l}$ is the distance between BS q_l and user u_k, and $10 \log 10(L_{u_k}^{q_l}) \sim N(0, 64)$. We also assume that $\sigma_{u_k}^2 = \sigma^2, \ \ \forall u_k \in \mathcal{U}$. All the simulation results are averaged over 100 channel realizations. The results shown for problem (2.8), (2.28) and (2.38) are those obtained after performing the de-biasing step.

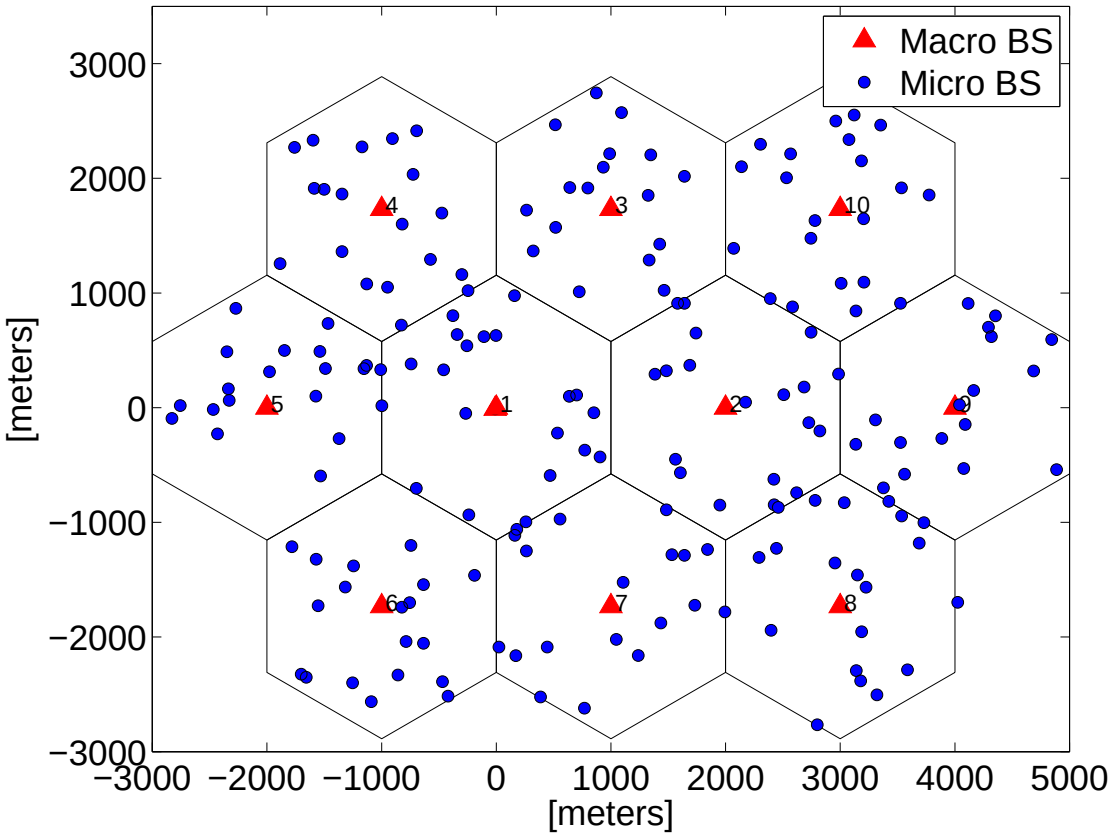


Figure 2.2: A randomly generated network configuration

The proposed algorithm is compared to the following two scenarios: 1) all the BSs are turned on; 2) in each cell, the central BS and a randomly selected fixed number of the remaining BSs are turned on. Note that for both of these cases, full JP is used within each cell. Clearly, the first scenario can serve as the performance upper bound, and the latter can serve as a reasonable heuristic algorithm to select active BSs since BSs and users are uniformly distributed in each cell.

In the first set of simulations, the total power minimization design criterion is considered. We set $U = 10$, $B = 20$, $M = 5$, and $\tau_{u_k} = 15\text{dB}$, $\forall u_k \in \mathcal{U}$. Furthermore, we assume that the power budget for BSs in the center of each cell is 10 dB while the budget for the rest of the BSs is set to be 5dB. Note that all the power considered here is relative to an implicit reference power. We apply the ADMM approach to solve the proposed formulation (2.8) with reweighting procedure. Since the objective QoS

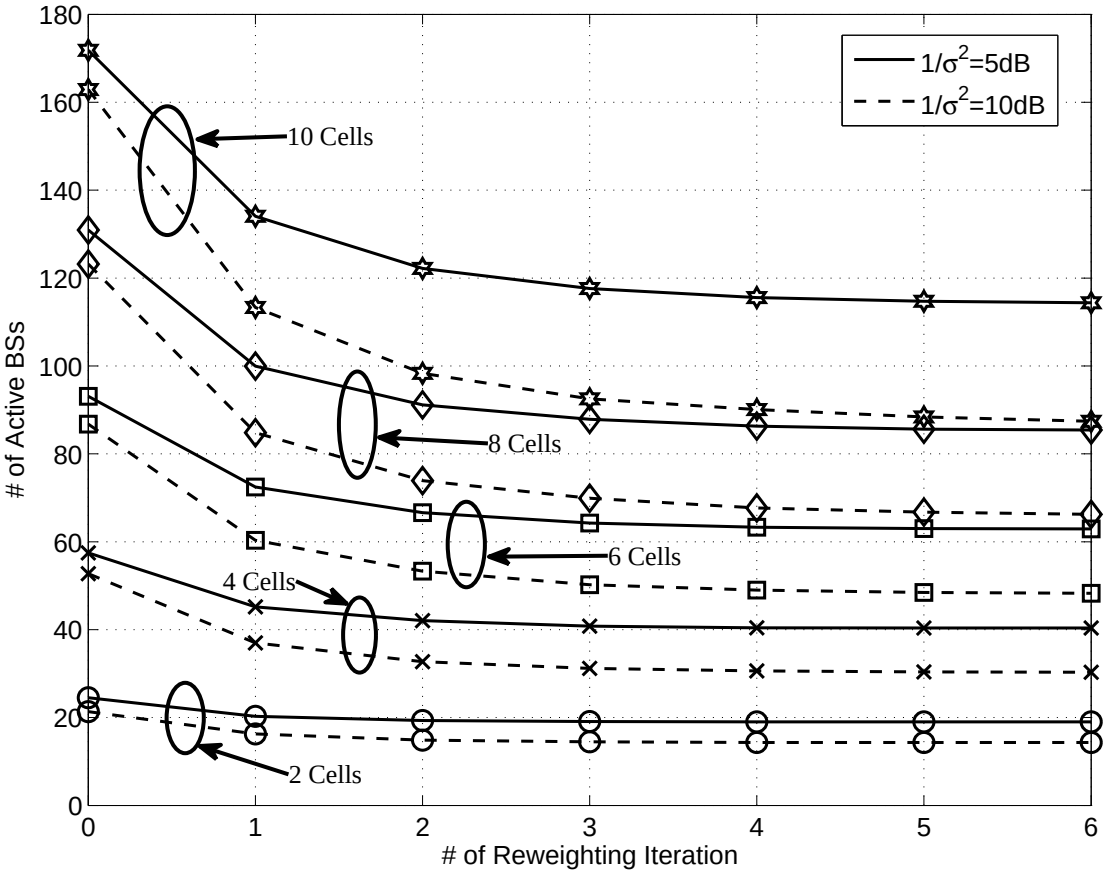


Figure 2.3: Number of active BSs after each reweighting procedure.

$\tau_{u_k}, \quad \forall u_k \in \mathcal{U}$ may not always be feasible, we declare that this realization is infeasible if a particular problem realization cannot converge within 2000 ADMM iterations. We select the stepsize as $\rho = 5$, and use the following stopping criterion

$$\max\left(\left\|\frac{\|vec(\mathbf{K})\|_\infty}{\max(1,\|\mathbf{K}\|_F)}\right\|, \left\|\frac{\mathbf{v}-\boldsymbol{w}}{\max(1,\|\mathbf{v}\|,\|\boldsymbol{w}\|)}\right\|_\infty, \max_{u_k\in\mathcal{U}}(|\kappa_{u_k}^2-\sigma^2|), \frac{f^{\min}(\boldsymbol{w}^{(t)})-f^{\min}(\boldsymbol{w}^{(t-1)})}{f^{\min}(\boldsymbol{w}^{(t-1)})}\right) < 10^{-4}.$$

In Fig. 2.3, we plot the number of active BSs after each reweighting procedure on $\beta_{b_k}, \quad \forall b_k \in \mathcal{B}$ for $1/\sigma^n = 5$dB and 10dB, respectively. From this figure, it can be observed that the number of active BSs decreases fast for the first 2 reweighting iterations, and converges within 6 reweighting iterations. In Fig. 2.4, the obtained minimum total

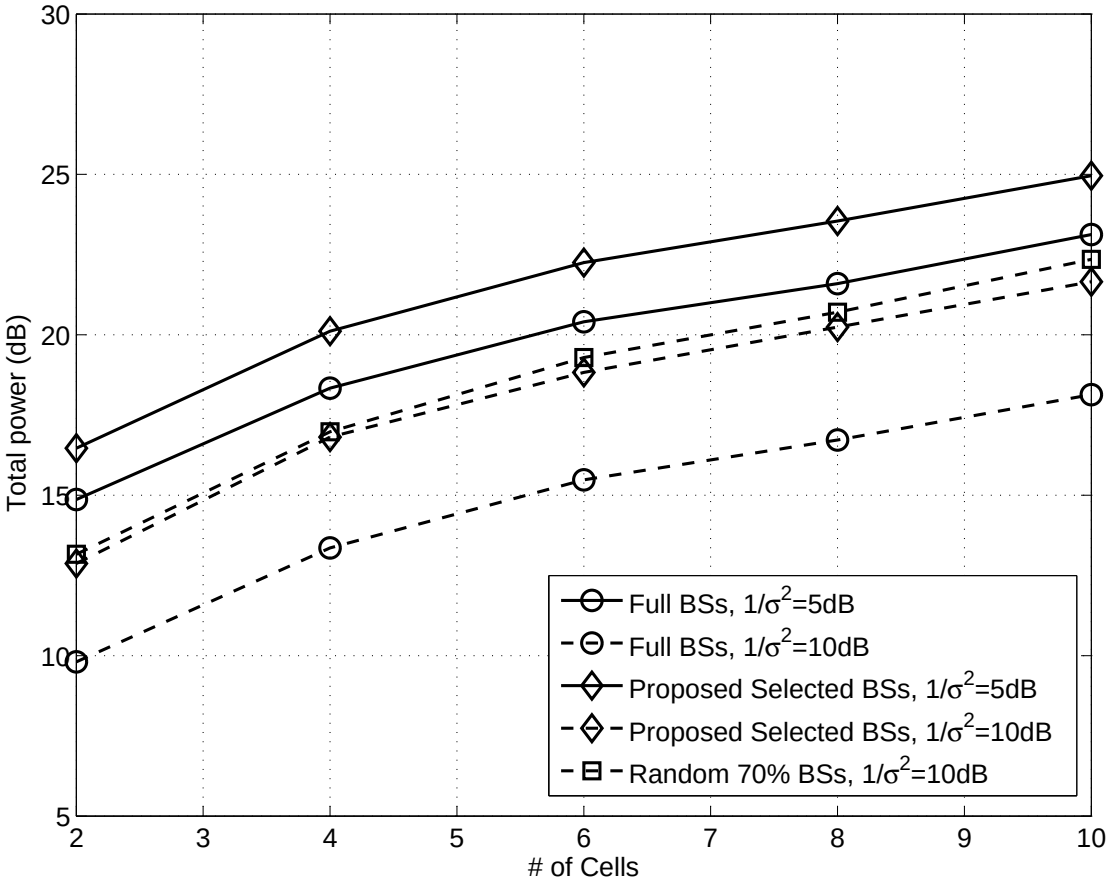


Figure 2.4: Power consumption for all scenarios considered.

power is plotted against the number of cells. We can observe that the minimum required power for BSs selected by the proposed formulation (2.8) is more than that achieved by activating all the BSs in each cell. This is reasonable since the latter serves as a lower bound of achievable power consumption. On the other hand, when $1/\sigma^2 = 10$dB, we compare the minimum power consumption achieved by the following two networks: i) a network with 70% of randomly activated BSs (the center BSs in each cell are always active) and ii) the network optimized by the proposed algorithm ($35.8\% \sim 43.45\%$ of BSs are activated for each number of cells). It can be observed that the proposed formulation is able to use much smaller number of BSs with similar total transmit power to support the same set of QoS constraints. This demonstrates the efficacy of the proposed method. Additionally, Fig. 2.5 plots the required number of ADMM iterations for the

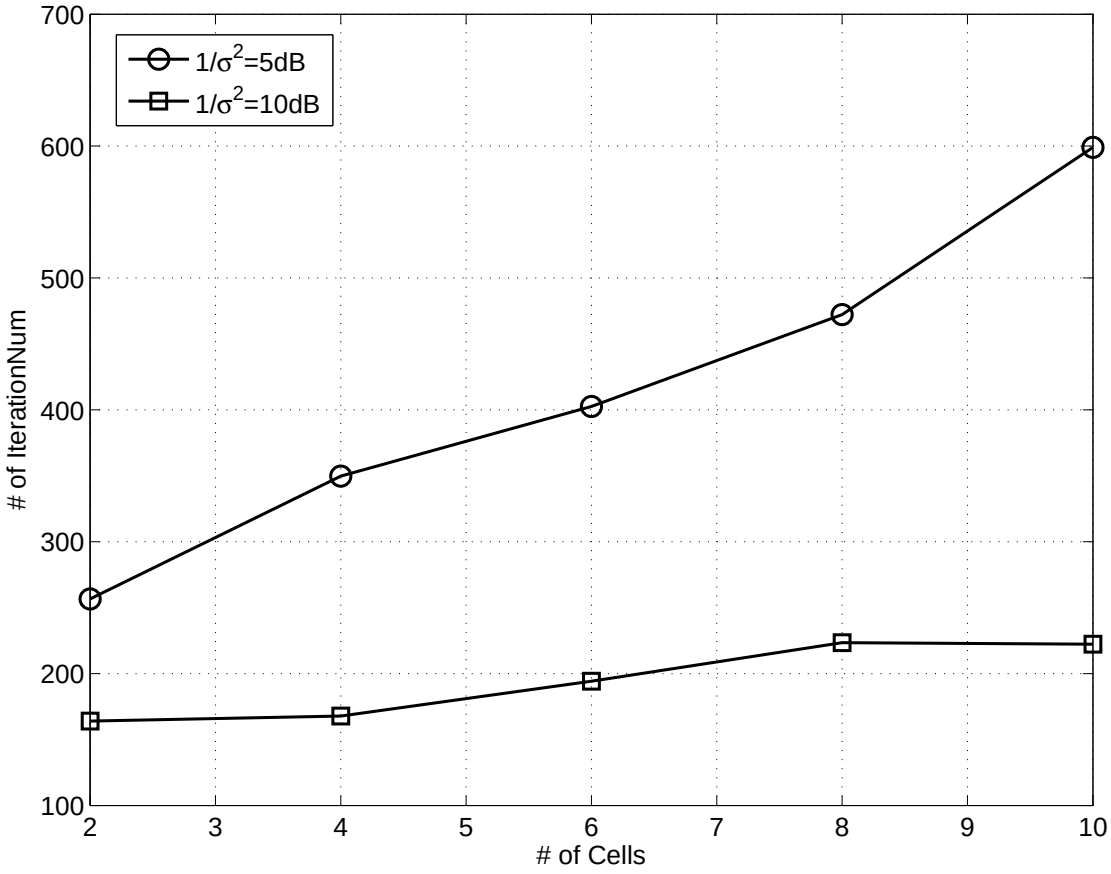


Figure 2.5: The required number of ADMM iterations for the scenario where all the BSs are active.

power minimization only design (2.4) (with all BSs being turned on). We observe that the proposed ADMM approach converges fairly fast. Note that the convergence speed depends on the channel quality, σ^2: when the channel condition is good enough, i.e., $1/\sigma^2 = 10$dB, it converges within 250 ADMM iterations.

In the second simulation set, the sum rate maximization design criterion is investigated. Let $U = 10$, $B = 10$, $N = 4$, $L = 2$ and P^{tot} denote the total power budget in each cell. The power budget for BSs located in the center of the cells is $P^{tot}/2$, and the rest of the BSs have equal power budgets. For simplicity, we set $\mu_{b_k} = \mu, \ \ \forall b_k \in \mathcal{B}$, $\lambda_k = \lambda, \ \ \forall k \in \mathcal{K}$, and $\sigma^2_{u_k} = 1, \ \ \forall u_k \in \mathcal{U}$. The reweighting procedure is performed until no BS reduction is possible or less than 50% of BSs is active. This is for fair comparison

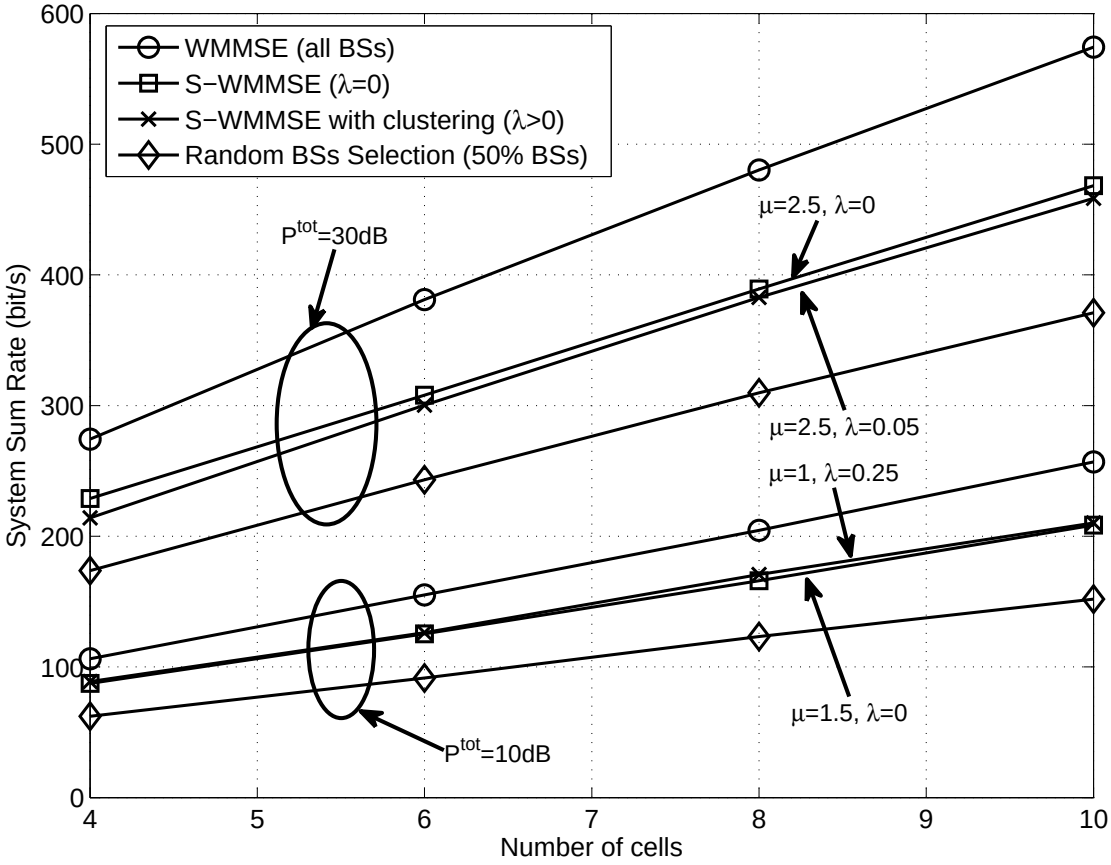
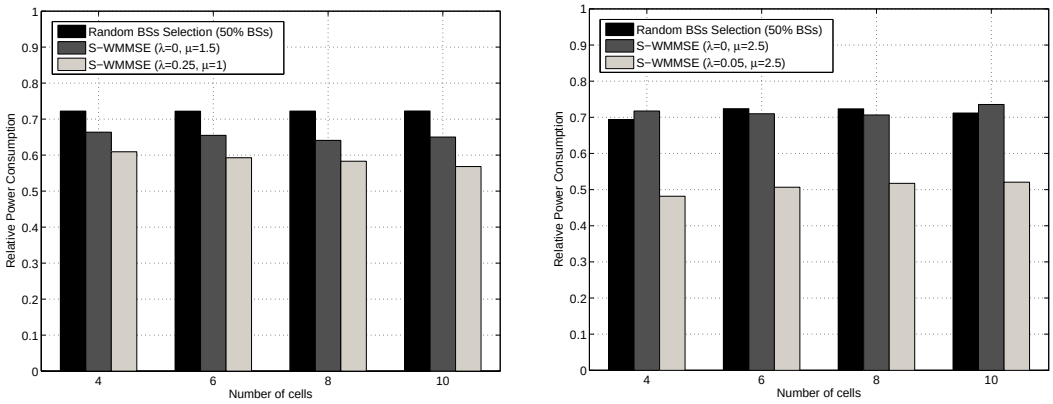


Figure 2.6: Sum rate performance comparison for different number of cells and total power budget.

with random selection scheme turning on 50% of BSs. In Fig. 2.6, the system sum rate performance for the proposed S-WMMSE algorithm is compared with $P^{tot} = 10$dB and 30dB. We can observe that S-WMMSE can achieve about 80% of the sum rate compared to the upper bound while activating around 50% BSs (see Tab. 2.2 for details about the number of active BSs). Furthermore, while the number of active BSs for S-WMMSE is about the same as the random selection scheme, the S-WMMSE can still achieve more than 34% and 23% improvement in sum rate performance for $P^{tot} = 10$dB and 30dB, respectively. It is worth noting that when BS clustering is considered, there is no sizable decrease in the sum rate performances. However, the total power consumption is significantly reduced; see Fig. 2.7. This is because when optimizing the BS clustering,

Number of Cells	4	6	8	10
WMMSE (all BSs)	40	60	80	100
Random BSs Selection (50% BSs)	20	30	40	50
S-WMMSE ($\mu = 1.5$, $\lambda = 0$), $P^{tot} = 10$dB	18.27	26.33	35.24	43.53
S-WMMSE ($\mu = 1$, $\lambda = 0.25$), $P^{tot} = 10$dB	20.18	28.67	38.51	47.04
S-WMMSE ($\mu = 2.5$, $\lambda = 0$), $P^{tot} = 30$dB	21.11	28.38	36.42	45.80
S-WMMSE ($\mu = 2.5$, $\lambda = 0.05$), $P^{tot} = 30$dB	20.21	28.73	37.80	46.95

Table 2.2: The number of active BSs v.s. different number of cells.



(a) $P^{tot} = 10$dB (b) $P^{tot} = 30$dB

Figure 2.7: Comparison of the power consumption for different schemes with varying P^{tot}. The total power used for the case where all BSs are active is normalized to 1.

the coverage of each BS is reduced, so does the interference level. As a result, less total transmit power is able to support similar sum rate performance.

In summary, our simulation results suggest that for the power minimization design criterion, the proposed ADMM approach can effectively reduce the BS usage while minimizing the required minimum power consumption. When considering the sum rate maximization design criterion, the proposed S-WMMSE algorithm can effectively reduce the BS usage and the size of BS cluster simultaneously. Finally, our simulations have only considered a simplified HetNet scenario: the uncorrelated Rayleigh channels with two types of BSs, for which the effectiveness of the proposed algorithms by selective shutdown of BSs is quite obvious. We caution that these results may not necessarily give

an accurate performance prediction for more realistic HetNet channel models, e.g., the channel models discussed in [118]. The latter will require further numerical experiments in the future.

Chapter 3

Min Flow Rate Maximization

In this chapter, we start to further consider a cloud-based heterogeneous network, i.e., C-RAN architecture, of base stations (BSs) connected via a backhaul network of routers and wired/wireless links with limited capacity. The optimal provision of such networks requires proper resource allocation across the radio access links in conjunction with appropriate traffic engineering within the backhaul network. In this chapter we propose an efficient algorithm for joint resource allocation across the wireless links and the flow control over the entire network. The proposed algorithm, which maximizes the min-rate among all the transmitted commodities, is based on a decomposition approach that leverages both the ADMM and the WMMSE algorithm. We show that this algorithm is easily parallelizable and converges globally to a stationary solution of the joint optimization problem. The proposed algorithm can also be extended to networks with multi-antenna nodes and other utility functions.

The organization of this chapter is as follows. In Sec. 4.2, the system model is introduced, and the considered joint optimization problem for C-RAN is formulated. In Sec. 3.2, the considered problem is investigated, and the corresponding proposed distributed algorithm is outlined with discussion on the distributed implementation issues. Moreover, the convergence analysis of the proposed algorithm is also provided. In Sec. 3.3, the closed-form solutions of each updating step is discussed in details. In Sec. 4.6, the efficacy and the efficiency of the proposed algorithms are demonstrated via extensive simulations.

3.1 System Model and Problem Formulation

We consider the downlink of a large-scale C-RAN illustrated in Fig. 1.1, where the data flows are from the network to the users. Let $\mathcal{V}$ denote the set of nodes in an C-RAN, which is comprised of a set of network routers $\mathcal{R}$, a set of BSs $\mathcal{B}$, and a set of mobile users $\mathcal{U}$. The BSs and mobile users are all equipped with one antenna. Let $\mathcal{L}$ denote the set of directed links that connect the nodes of $\mathcal{V}$. In the ensuing sections, the mth data flow demanded by the destination node $d_m \in \mathcal{V}$ from the source node $s_m \in \mathcal{V}$ is called the *commodity* m. We assume there are a total of M commodities to be transported over the network. With this definition, a mobile user $u \in \mathcal{U}$ can serve as the destination nodes for more than one commodity. For each commodity m, $r_m \geq 0$ denotes its flow rate, and $f_{l,m} \geq 0$ denotes its rate on link $l \in \mathcal{L}$.

The C-RAN has a set of directed links $\mathcal{L}$ consisting of both wired and wireless links. The wired links connect routers in $\mathcal{R}$ and BSs in $\mathcal{B}$, and is denoted as $\mathcal{L}^w \triangleq \{(s_l, d_l) \in \mathcal{L} \mid \forall\, s_l, d_l \in \mathcal{R} \cup \mathcal{B}\}$. Here (s_l, d_l) denotes the directed link from node s_l to node d_l. The wireless links provide single-hop connections between the BSs and the mobile users. We assume that each BS divides the spectrum into F orthogonal frequency subchannels, and refer to each subchannel as a *wireless link*. Thus a source node, a destination node and a subchannel uniquely define a wireless link. The set of wireless links can then be represented as $\mathcal{L}^{wl} \triangleq \{(s_l, d_l, f_l) \in \mathcal{L} \mid \forall\ s_l \in \mathcal{B},\ \forall\ d_l \in \mathcal{U},\ f_l = 1 \sim F\}$ with $l = (s_l, d_l, f_l)$ being the wireless link from BS s_l to mobile user d_l on subchannel f_l. For $l,\ n \in \mathcal{L}^{wl}$, $l \neq n$, the channel tap from BS s_n to mobile user d_l via subchannel f_l is denoted as $h_{ln} \in \mathbb{C}$. It is nonzero if either the links l and n occupy the same frequency subchannel (i.e., $f_l = f_n$), or s_l and d_n are not too far away from each other. Using this notation, the wireless link l is said to be interfered by the set of wireless links $I(l) \triangleq \{n \in \mathcal{L}^{wl} \mid h_{ln} \neq 0\}$. Note that $l \in I(l)$ by definition.

Next we introduce a few system level constraints in both the backhaul and the access networks. The first set of constraints is related to the link capacity. Assume each wired link $l \in \mathcal{L}^w$ has a fixed capacity, C_l. The total flow rate on link l is constrained by

$$\sum_{m=1}^{M} f_{l,m} = \mathbf{1}^T \mathbf{f}_l \leq C_l,\ \forall\, l \in \mathcal{L}^w. \tag{3.1}$$

where $\mathbf{1}$ is the all-one vector and $\mathbf{f}_l \triangleq [f_{l,1}, \ldots, f_{l,M}]^T$. For a wireless link l, BS s_l allocates power p_l for mobile user d_l. Assume a linear precoder v_l is used by each BS, then we have $v_l = \sqrt{p_l} \in \mathbb{R}$. Letting v_l be a real number is without loss of generality, because single antenna is used by each BS and each user. The advantage of using transmit precoders as design variables is that it facilitates the subsequent algorithm development and analysis, and allows easy extension to MIMO networks. Assume that each mobile user d_l treats the interference from interfering links $I(l) \setminus \{l\}$ as noise, then the total flow rate constraint on the wireless link $l \in \mathcal{L}^{wl}$ is

$$\sum_{m=1}^{M} f_{l,m} = \mathbf{1}^T \mathbf{f}_l \leq \bar{f}_l(\mathbf{v}) \triangleq \log\left(1 + \frac{|h_{ll}|^2 v_l^2}{\sum\limits_{n \in I(l)\setminus\{l\}} |h_{ln}|^2 v_n^2 + \sigma_{d_l}^2}\right), \ \forall\, l \in \mathcal{L}^{wl}, \tag{3.2}$$

where $\mathbf{v} \triangleq [\{v_l\}_{\forall\, l \in \mathcal{L}^{wl}}]^T$; $\bar{f}_l(\mathbf{v})$ is the achievable rate on the wireless link l for a given precoders $\mathbf{v}$; and $\sigma_{d_l}^2$ is the variance of AWGN at mobile user d_l.

The second set of constrains has to do with the per-node flow conservation constraint. That is, for any node $v \in \mathcal{V}$, the total incoming flow should be equal to the total outgoing flow:

$$\sum_{l \in \mathrm{In}(v)} f_{l,m} + 1_{v=s_m} r_m = \sum_{l \in \mathrm{Out}(v)} f_{l,m} + 1_{v=d_m} r_m, \quad m = 1 \sim M, \ \forall\, v \in \mathcal{V}, \tag{3.3}$$

where $\mathrm{In}(v) \triangleq \{l \in \mathcal{L} \mid \ d_l = v\}$ and $\mathrm{Out}(v) \triangleq \{l \in \mathcal{L} \mid \ s_l = v\}$ denote the set of links going into and coming out of a node v respectively.

The third set of constraints requires that the transmit power used by each BS $s \in \mathcal{B}$ should be less than a given budget $P_s \geq 0$:

$$\sum_{l \in \mathrm{Out}(s) \bigcap \mathcal{L}^{wl}} v_l^2 \leq P_s, \ \forall\, s \in \mathcal{B}. \tag{3.4}$$

In this work, we are interested in maximizing the minimum flow rate of all commodities, $r_{\min}$, while jointly performing the following tasks *1)*: route M commodities from node s_m to node d_m, $m = 1 \sim M$; and *2)* design the linear precoder at each BS. This problem can be formulated as

$$\max_{\mathbf{v},\, \mathbf{r}} \quad r_{\min} \tag{3.5a}$$

$$\text{s.t.} \quad \mathbf{r} \geq 0, \ r_m \geq r_{\min}, \ m = 1 \sim M \tag{3.5b}$$

$$(3.1), \ (3.2), \ (3.3), \ \text{and} \ (3.4), \tag{3.5c}$$

where $\mathbf{r} \triangleq [r_{\min}, \{r_m, f_{l,m} \mid l \in \mathcal{L}\}_{m=1\sim M}]^T$. The constraint (3.5b) is due to the non-negativeness of flow rates, and the fact that $r_{\min}$ is the min-rate of all commodities. Optimizing the min-rate utility results in a fair rate allocation, and such utility has been adopted by many recent works in both the SDN and wireless communities; see [64, 119] and the references therein.

Remark 3 *(Difficulties of Solving Problem* (3.5)*) Problem* (3.5) *has several distinctive structures:*

i) The feasible set of (3.5) *is nonconvex as a result of the wireless rate constraints* (3.2)*.*

ii) The design variables $\mathbf{v}$ *and* $\mathbf{r}$ *are tightly coupled through the rate expressions in the constraints* (3.2) *and* (3.3)*.*

iii) The size of the problem is very large.

Together, these special structures make the existing techniques for min-rate maximization [27, 60–62] inapplicable. For example, [27, 60–62] exploit the structure of signal-to-interference-plus-noise ratio of wireless links, but here we need to directly deal with the users' rates $\bar{f}_l(\mathbf{v})$ *as they are coupled in the constraints. Moreover, the size of the problem makes it computationally very expensive to repeatedly solve the problems formulated in [27, 61, 62] using standard solvers. Hence, the structure of signal-to-interference-plus-noise ratio of wireless links exploited by existing works cannot be applied. The wireless rate function,* $\bar{f}_l(\mathbf{v})$ *should be dealt directly.*

Remark 4 *(Dynamic BS selection) By solving problem* (3.5)*, a subset of BSs is dynamically selected to serve each user. Specifically, for commodity* m*, it is possible for two different wireless links* $l \neq n$ *with* $d_l = d_n$ *to each carry part of commodity* m*, i.e.,* $f_{l,m} > 0$ *and* $f_{n,m} > 0$*. Allowing BS cooperation is an important feature of the envisioned next generation cellular networks [11, 14]. Our proposed formulation and algorithm can be extended to incorporate more advanced cooperation schemes, e.g., joint processing between BSs, possibly by using different flow control mechanisms such as network coding [80]. For the envisioned next generation networks, which relies heavily on*

various BS cooperation schemes to improve the transmission rate, the proposed formulation accompanied with the algorithm can be extended with given BS cluster for each commodity and/or applying network coding in the backhaul network [80].

In the following, we propose an efficient distributed algorithm to compute a stationary solution of the problem (3.5).

3.2 Joint Traffic Engineering and Interference Management

In this section, we propose a distributed algorithm that solves problem (3.5) to a stationary solution. We emphasize that the difficulty of this problem comes from the nonconvexity in the wireless link flow rate constraints (3.2), as well as the way that the flow rates are coupled in the flow conservation constraints (3.3).

3.2.1 Algorithm Outline

We propose to integrate two existing algorithms to solve problem (3.5). The first one is the max-min WMMSE algorithm developed in [64], which is used for min-rate maximization. The second is the well-known ADMM algorithm for large-scale optimization. Central to the proposed approach is the utilization of a relationship between achievable flow rate for wireless link $l \in \mathcal{L}^{wl}$, i.e., $\bar{f}_l(\mathbf{v})$, and the mean square error (MSE) for estimating the message transmitted on link l. Let us use $e_l(u_l, \mathbf{v})$ to denote the MSE on link l when user d_l applies a linear receive coefficient $u_l \in \mathbb{R}$ to decode the message. Then $e_l(u_l, \mathbf{v})$ is given by

$$e_l(u_l, \mathbf{v}) = (1 - u_l|h_{ll}|v_l)^2 + u_l^2 \left(\sum_{n \in I(l) \setminus \{l\}} |h_{ln}|^2 v_n^2 + \sigma_{d_l}^2 \right).$$

The following rate-MSE relationship, a specialization of the results developed in [64, Lemma 3] to the single antenna scenario, is a key property used in our subsequent algorithm design.

Lemma 2 *[64, Lemma3] For a given $l \in \mathcal{L}^{wl}$, $\bar{f}_l(\mathbf{v})$ can be equivalently expressed as*

$$\bar{f}_l(\mathbf{v}) = \max_{u_l,\, w_l} \left[1 + \log(w_l) - w_l e_l(u_l, \mathbf{v})\right], \tag{3.6}$$

where $w_l > 0$ is the scalar weight of MSE on link l.

Lemma 2 reformulates $\bar{f}_l(\mathbf{v})$ by introducing two extra sets of variables $\mathbf{u} \triangleq \{u_l \mid l \in \mathcal{L}^{wl}\}$ and $\mathbf{w} \triangleq \{w_l \mid l \in \mathcal{L}^{wl}\}$, with one pair of variables $\{u_l, w_l\}$ for each wireless link l. Hence, we reformulate problem (3.5) by replacing $\bar{f}_l(\mathbf{v})$ with its weighted MSE. We call such new constraint a *rate-MSE constraint*. Using this relationship, we consider the following modified version of problem (3.5), with two additional variable sets $\mathbf{u}$ and $\mathbf{w}$:

$$\max_{\mathbf{v},\, \mathbf{r},\, \mathbf{u},\, \boldsymbol{w}} \; r_{\min} \tag{3.7a}$$

$$\text{s.t. (3.1), (3.3), (3.4), and (3.5b)} \tag{3.7b}$$

$$\mathbf{1}^T \mathbf{f}_l \le c_{1,l} + c_{2,l} v_l - \sum_{n \in I(l)} c_{3,ln} v_n^2, \; \forall\, l \in \mathcal{L}^{wl}, \tag{3.7c}$$

where $(c_{1,l}, c_{2,l}, c_{3,ln})$ are given by $c_{1,l} = 1 + \log(w_l) - w_l(1 + \sigma_{d_l}^2 u_l^2)$, $c_{2,l} = 2 w_l u_l |h_{ll}|$, and $c_{3,ln} = w_l u_l^2 |h_{ln}|^2$.

Why do we include these extra optimization variables $\mathbf{u}$ and $\mathbf{w}$? First we observe that for any given $\{\mathbf{r}, \mathbf{v}\}$, the optimal $\mathbf{u}$ (resp. $\mathbf{w}$) for (3.6) can be obtained while $\mathbf{w}$ (resp. $\mathbf{u}$) is held fixed. Moreover, these optimal solutions can be expressed in closed form for any $l \in \mathcal{L}^{wl}$:

$$u_l(\mathbf{v}) = \left(\sum_{n \in I(l)} |\, h_{ln}|^2 v_n^2 + \sigma_{d_l}^2 \right)^{-1} |h_{ll}| v_l, \tag{3.8}$$

$$w_l(\mathbf{v}) = \left(1 - |h_{ll}| v_l u_l(\mathbf{v})\right)^{-1}. \tag{3.9}$$

These expressions suggest that for each wireless link l, the variables u_l and w_l can be updated locally at mobile user d_l, which is independent to other mobile users, if the interference plus noise and local channel state information are locally known to the users. Moreover, when $\mathbf{u}$ and $\mathbf{w}$ are fixed, the problem for updating $\{\mathbf{r}, \mathbf{v}\}$ is convex (note that (3.7) is a convex quadratic problem on the precoders $\mathbf{v}$) and can be solved in polynomial time. Hence, we propose to apply the alternating optimization technique to

solve problem (3.7); see the N-MaxMin Algorithm in Table 3.1 for a detailed description. The following is our main convergence result. Its proof is relegated to Appendix C.

Theorem 4 *The sequence* $\{\mathbf{r}^{(t)}, \mathbf{v}^{(t)}\}$ *generated by the N-MaxMin Algorithm converges to a stationary solution of problem* (3.5)*. Moreover, every global optimal solution of problem* (3.5) *corresponds to a global optimal solution of the reformulated problem* (3.7)*, and they achieve the same objective value.*

Remark 5 *(Relationship to [64, Theorem 2]) The N-MaxMin Algorithm leverages the rate-MSE relationship developed in [64] to deal with the nonconvex constraint* (3.2) *in the general setting of C-RAN, and at the same time uses the ADMM (algorithm 2) to determine the backhaul network flow. The above convergence result generalizes the one developed in [64] due to the new network flow constraints* (3.1) *and* (3.3) *involved. Similar result has recently been proposed [120], but only the simplified single-hop backhaul network is considered.*

Remark 6 *(Extension to Multiple Antenna Case and Other Utility Functions) The proposed algorithm can easily handle nodes with multiple transmit/receive antennas. Specifically, we can use the matrix version of rate-MSE relationship (see [64, Lemma 3]) to replace the capacity constraint on each wireless link. Moreover, the convergence proof uses the fact that, at optimality, at least one of the constraints* $r_m \geq r_{\min}$*,* $m = 1 \sim M$*, is active. For other utility functions, e.g., proportional fairness* $\sum_{m=1}^{M} \log(r_m) \geq r_{\text{proportional}}$*, this property still holds, so the convergence analysis can be extended to these other cases.*

3.2.2 An ADMM Approach for Updating $\{\mathbf{r}, \mathbf{v}\}$

Unlike the computation of $\mathbf{u}$ and $\mathbf{w}$, the updates for $\{\mathbf{r}, \mathbf{v}\}$ do not have closed forms. This can be a problem for large networks as the size of the subproblem can be huge. Below we propose to use the ADMM algorithm as a subroutine to update $\{\mathbf{r}, \mathbf{v}\}$. We choose ADMM because it is well-suited for distributed and parallel implementation, which is attractive for the considered C-RAN.

To apply ADMM, we first reformulate the subproblem for $\{\mathbf{r}, \mathbf{v}\}$ into the *standard form* (cf. (A.1)). To this end, we appropriately split the variables in the coupling

Network Max-Min WMMSE (N-MaxMin) Algorithm:

1: **Initialization** Generate a feasible set of variables $\{\mathbf{r}, \mathbf{v}\}$, and let $t = 1$.
2: **Repeat**
3: $\quad \mathbf{u}^{(t)}$ is updated by (3.8)
4: $\quad \mathbf{w}^{(t)}$ is updated by (3.9)
5: $\quad \{\mathbf{r}^{(t)}, \mathbf{v}^{(t)}\}$ is updated by solving the problem (3.7) via Algorithm 2 in Table 3.2
6: $\quad t = t + 1$
7: **Until** A desired stopping criteria is met

Table 3.1: Network Max-Min WMMSE (N-MaxMin) Algorithm

constraints (3.3) and (3.7c) for $\mathbf{r}$ and $\mathbf{v}$. Then we show that each step of the resulting algorithm is easily computable and amenable for distributed implementation.

We first observe that each flow rate $f_{l,m}$ is shared among *two* flow conservation constraints, one for node s_l and the other for node d_l. To induce separable subproblems and enable distributed computation, two local auxiliary copies of $f_{l,m}$ are introduced, namely $\hat{f}_{l,m}^{s_l}$ and $\hat{f}_{l,m}^{d_l}$, and they are, respectively, stored at node s_l and node d_l. Similarly, we introduce two local auxiliary copies for each commodity rate, denoted as $\hat{r}_m^{s_m}$, $\hat{r}_m^{d_m}$, $m = 1 \sim M$, and store them at the source and the destination node of each commodity, respectively. That is, we introduce the following auxiliary variables:

$$\hat{f}_{l,m}^{s_l} = f_{l,m},\ \hat{f}_{l,m}^{d_l} = f_{l,m},\ \forall\, l \in \mathcal{L},\ \forall\, m; \tag{3.10a}$$

$$\hat{r}_m^{s_m} = r_m,\ \hat{r}_m^{d_m} = r_m,\ \forall\, m. \tag{3.10b}$$

The flow rate conservation constraints on node $v \in \mathcal{V}$ can then be rewritten as

$$\sum_{l \in \mathrm{In}(v)} \hat{f}_{l,m}^{v} + 1_{v=s_m} \hat{r}_m^{v} = \sum_{l \in \mathrm{Out}(v)} \hat{f}_{l,m}^{v} + 1_{v=d_m} \hat{r}_m^{v},\ m = 1 \sim M. \tag{3.11}$$

Moreover, for the rate-MSE constraint (3.7c), we introduce several copies of the transmit precoder on a given wireless link $n \in \mathcal{L}^{wl}$, i.e.

$$v_{ln} = v_n,\ \forall\, l \text{ s.t. } n \in I(l). \tag{3.12}$$

Intuitively, by doing such variable splitting, each variable v_{nl} will only appear in *a single* rate-MSE constraint. For a given link $l \in \mathcal{L}^{wl}$, its rate-MSE constraint only

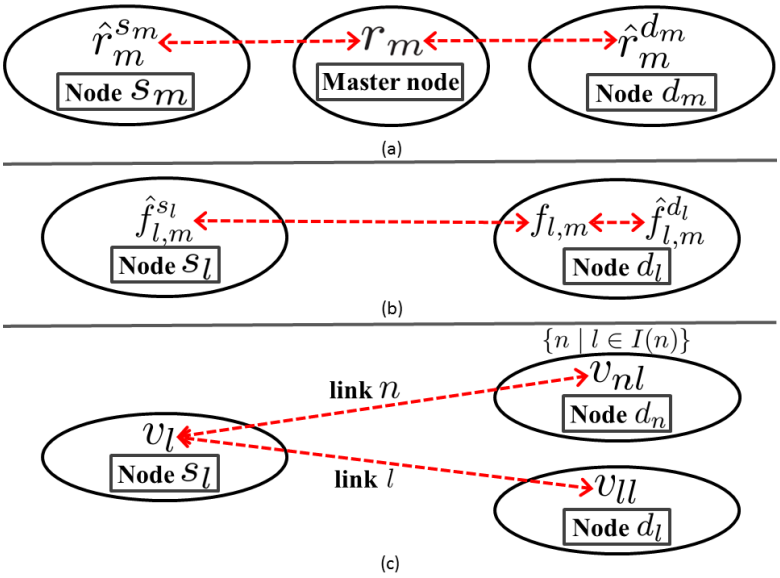


Figure 3.1: The structure of the design variables and the introduced auxiliary variables. The variables connected by dash lines should be equal to each other. (a) For the mth commodity, $m = 1 \sim M$. (b) For the wired link $l \in \mathcal{L}^w$. (c) For the wireless link $l \in \mathcal{L}^{wl}$.

depends on the set of precoders $\{v_{ln} \mid \forall\, n \in I(l)\}$, as can be seen below

$$\mathbf{1}^T \mathbf{f}_l \le c_{1,l} + c_{2,l} v_{ll} - \sum_{n \in I(l)} c_{3,ln} v_{ln}^2, \ \forall\, l \in \mathcal{L}^{wl}. \tag{3.13}$$

In addition, to facilitate the analysis of the convergence, another auxiliary variable $\hat{r}$ is introduced such that $r_{\min} = \hat{r}_{\min}$. The relationship between the design variables and the introduced auxiliary variables is illustrated in Fig. 3.1.

Using these new variables, the updating step for $\{\mathbf{r}, \mathbf{v}\}$ is equivalently expressed as

$$\begin{aligned} &\max \ (r_{\min} + \hat{r}_{\min})/2 \\ &\text{s.t. } (3.1),\ (3.4),\ (3.5b),\ (3.10),\ (3.11),\ (3.12),\ (3.13),\ \text{and } r_{\min} = \hat{r}_{\min}. \end{aligned} \tag{3.14}$$

It is important to note that the constraints of problem (3.14) (except the linear equality constraints $r_{\min} = \hat{r}_{\min}$, (3.10) and (3.12)) are now separable between two optimization variable sets

1. The tuple $\{\mathbf{r}, \hat{\mathbf{v}}\}$ where $\hat{\mathbf{v}} \triangleq [\,[v_{nl}, \forall\, n \in \bar{I}(l)], \forall\, l \in \mathcal{L}^{wl}]^T$ with $\bar{I}(l) \triangleq \{n \mid l \in I(n)\}$ being the set of wireless links with which l interferes.

2. The tuple $\{\hat{\mathbf{r}}, \mathbf{v}\}$ where $\hat{\mathbf{r}} \triangleq [\ \hat{r}_{\min}, \{\hat{r}_m^{s_m}, \hat{f}_{l,m}^{s_l} \mid l \in \mathcal{L}\}_{m=1}^{M}, \{\hat{r}_m^{d_m}, \hat{f}_{l,m}^{d_l} \mid l \in \mathcal{L}\}_{m=1}^{M}\ ]^T$.

Furthermore, we can write the linear equalities $r_{\min} = \hat{r}_{\min}$, (3.10) and (3.12) as $\mathbf{Cr} = \hat{\mathbf{r}}$, $\mathbf{Dv} = \hat{\mathbf{v}}$ with

$$\mathbf{C} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & \mathbf{I} & \mathbf{0} & \mathbf{I} & \mathbf{0} \\ 0 & \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{I} \end{bmatrix}^T \ ; \ \mathbf{D} = \text{blkdg}[\{\mathbf{1}_l\}_{l \in \mathcal{L}^{wl}}],$$

where $\mathbf{1}_l$ is an all one column vector of size equal to $|\bar{I}(l)|$. Based on this compact representation, let us use δ, $\{\delta_{l,m}^{s_l}, \delta_{l,m}^{d_l}\}$, $\{\delta_m^{s_m}, \delta_m^{d_m}\}$, and $\{\theta_{nl}\}$ to denote the Lagrange multipliers for equality constraints $r_{\min} = \hat{r}_{\min}$, (3.10a), (3.10b), and (3.12), respectively. Collect these multipliers to form vectors $\boldsymbol{\delta}$ and $\boldsymbol{\theta}$. Let $\rho_1 > 0$ and $\rho_2 > 0$ denote some constants.

Then the partial augmented Lagrangian function for problem (3.14) is given by

$$\begin{aligned} &L_{\rho_1,\rho_2}(\mathbf{r}, \hat{\mathbf{v}}, \hat{\mathbf{r}}, \mathbf{v}; \boldsymbol{\delta}, \boldsymbol{\theta}) \\ &= \frac{r_{\min} + \hat{r}_{\min}}{2} + \underbrace{\left[\boldsymbol{\delta}^T(\hat{\mathbf{r}} - \mathbf{Cr}) - \frac{\rho_1}{2}\|\hat{\mathbf{r}} - \mathbf{Cr}\|^2\right]}_{\text{enforcing linear constraints (3.10)}} + \underbrace{\left[\boldsymbol{\theta}^T(\mathbf{Dv} - \hat{\mathbf{v}}) - \frac{\rho_2}{2}\|\mathbf{Dv} - \hat{\mathbf{v}}\|^2\right]}_{\text{enforcing linear constraints (3.12)}}. \end{aligned} \tag{3.15}$$

Now it is clear that the ADMM can be used to solve (3.14). The resulting algorithm, described in Table 3.2, is referred to as Algorithm 2. The convergence of this algorithm to the optimal solutions of problem (3.14) is readily implied by Proposition 1 in Appendix A (note that both $\mathbf{C}^T\mathbf{C}$ and $\mathbf{D}^T\mathbf{D}$ are full rank matrices).

The detailed step-by-step specification of Algorithm 2 is given in Section 3.3. The main feature from the derivation therein is that each step in Algorithm 2 can be computed distributedly in closed-form. More specifically, let there be a master node to coordinate the flow rate for all commodities, then the terms of partial augmented Lagrangian function (3.15) are separable across each link and node. Similarly, the constraints of Step 3 and 4 in Table 3.2 are also separable across links and nodes, respectively. Given these facts, Step 3 of the algorithm is decomposable *among all links* in the system. Step 4 of the algorithm is decomposable *among all the nodes* in the system. Moreover, the update for each commodity can be done independently at each node. For the dual

Algorithm 2: ADMM for (3.14):

1: **Initialize** all primal variables $\mathbf{r}^{(0)}$, $\hat{\mathbf{r}}^{(0)}$, $\mathbf{v}^{(0)}$, $\hat{\mathbf{v}}^{(0)}$ (not necessarily a feasible solution for (3.14)), and all dual variables $\boldsymbol{\delta}^{(0)}$, $\boldsymbol{\theta}^{(0)}$; set $t = 0$

2: **Repeat**

3: Solve the following problem and obtain $\mathbf{r}^{(t+1)}, \hat{\mathbf{v}}^{(t+1)}$:

$$\max_{\mathbf{r},\, \hat{\mathbf{v}}} \; L_{\rho_1,\rho_2}(\mathbf{r}, \hat{\mathbf{v}}, \hat{\mathbf{r}}^{(t)}, \mathbf{v}^{(t)}; \boldsymbol{\delta}^{(t)}, \boldsymbol{\theta}^{(t)})$$
$$\text{s.t. (3.1), (3.5b), and (3.13).} \tag{3.16}$$

This step can be *solved in parallel across all links*, cf. (3.20), (3.22), and (3.24).

4: Solve the following problem and obtain $\hat{\mathbf{r}}^{(t+1)}, \mathbf{v}^{(t+1)}$:

$$\max_{\hat{\mathbf{r}},\, \mathbf{v}} \; L_{\rho_1,\rho_2}(\mathbf{r}^{(t+1)}, \hat{\mathbf{v}}^{(t+1)}, \hat{\mathbf{r}}, \mathbf{v}; \boldsymbol{\delta}^{(t)}, \boldsymbol{\theta}^{(t)})$$
$$\text{s.t. (3.4) and (3.11).} \tag{3.17}$$

This problem can be *solved in parallel across all nodes*, cf. (3.25), (3.27), and (3.29).

5: Update the Lagrange dual multipliers $\boldsymbol{\delta}^{(t+1)}$ and $\boldsymbol{\theta}^{(t+1)}$ by

$$\begin{aligned} \boldsymbol{\delta}^{(t+1)} &= \boldsymbol{\delta}^{(t)} - \rho_1(\hat{\mathbf{r}}^{(t+1)} - \mathbf{C}\mathbf{r}^{(t+1)}), \\ \boldsymbol{\theta}^{(t+1)} &= \boldsymbol{\theta}^{(t)} - \rho_2(\mathbf{D}\mathbf{v}^{(t+1)} - \hat{\mathbf{v}}^{(t+1)}). \end{aligned} \tag{3.18}$$

6: $t = t + 1$

7: **Until** Desired stopping criterion is met

Table 3.2: Summary of the proposed Algorithm 2

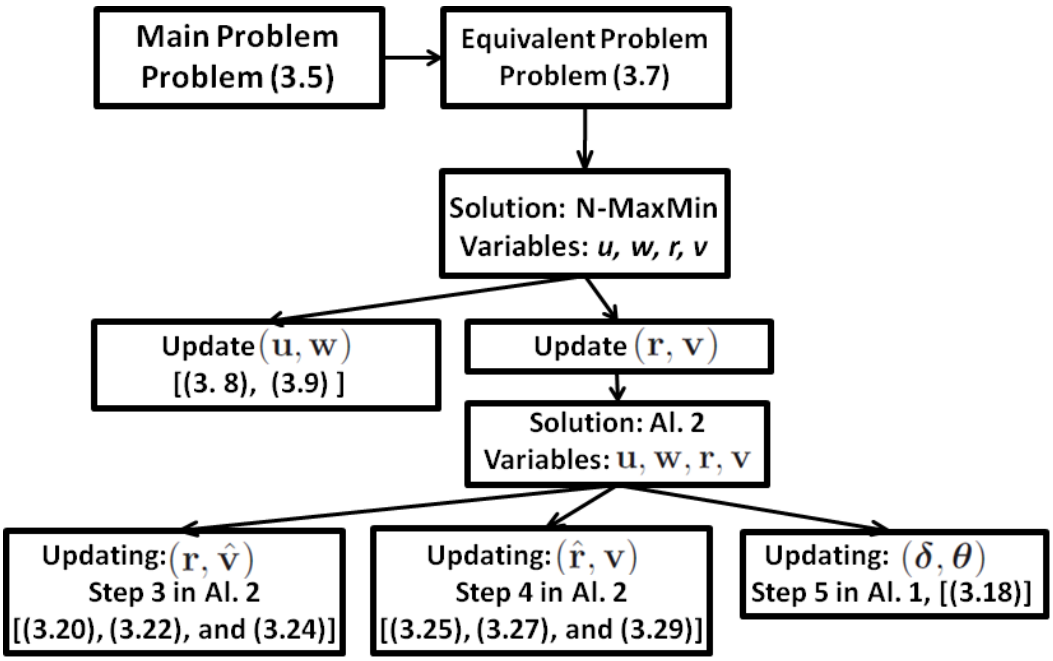


Figure 3.2: Flow chart of the proposed solution approach (3.5).

updates (3.18), it can be done in each network node independently. These properties allow the entire algorithm to be easily implemented in a parallel fashion, and for each ADMM iteration, the computation complexity is $O(M|\mathcal{V}| + |\mathcal{L}|)$. Fig. 3.2 provides a flow chart showing the relationship among different subroutines.

3.2.3 Necessary Information Exchange

In this subsection, we elaborate how the N-MaxMin algorithm can be implemented distributedly.

Let us first look at the implementation for the backhaul network (i.e., ignoring the wireless links). Consider the update of the optimization variable $\mathbf{r}$ in Step 3 of Algorithm 1. In this step, to update $\{r_{\min}, \{r_m\}_{m=1}^M\}$ (cf. Step (i) in Appendix 3.3.1), node $v_m \in \{s_m, d_m\}$ of each commodity m should respectively send $\left(\hat{r}_m^{v_m} - \frac{\delta_m^{v_m}}{\rho_1}\right)$ to the assumed master node. After the master node updates $\{r_{\min}, \{r_m\}_{m=1}^M\}$, it would transmit r_m back to these nodes. For updating $\{\mathbf{f}_l \mid \forall\, l \in \mathcal{L}^w\}$ (cf. step (ii) in Section

3.3.1), the procedure is decoupled across *each link*. Without loss of generality, we can let the destination node of each link $l \in \mathcal{L}^w$, d_l, perform the bisection updating step. The source node s_l should transmit M real values, $\{\hat{f}_{l,m}^{s_l} - \frac{\delta_{l,m}^{s_l}}{\rho_1}\}_{m=1}^M$, to d_l. After updating $\mathbf{f}_l$, d_l would transmit them back to s_l. After $\mathbf{r}$ is computed, the second block variables $\hat{\mathbf{r}}$ and the Lagrangian dual variable $\boldsymbol{\delta}$ are updated in each node without any additional information exchange, cf. step (ii) in Section 3.3.2 and (3.18).

Next we discuss the implementation for the wireless part, i.e., the update for $\mathbf{v}$, $\hat{\mathbf{v}}$, and the wireless links of $\mathbf{r}$ and $\hat{\mathbf{r}}$. We assume that for each wireless link $l \in \mathcal{L}^{wl}$ *i)* mobile user d_l has local channel state information from all interfering BSs, i.e., h_{ln}, $\forall\, n \in I(l)$; and *ii)* u_l and w_l are updated at the mobile user side. Hence $(c_{1,l}, c_{2,l}, c_{3,ln})$ are known locally at mobile user d_l. Let us first look at the update for $\hat{\mathbf{v}} \cup \{\mathbf{f}_l \mid \forall\, l \in \mathcal{L}^{wl}\}$ (cf. step (iii) in Section 3.3.1). Recall that this step is decoupled across wireless links, and all necessary information needed for the computation (such as $\mathbf{u}$, $\mathbf{w}$, $\mathbf{v}$ and the channel state information) is available at each user except $\{\hat{f}_{l,m}^{s_l} - \frac{\delta_{l,m}^{s_l}}{\rho_1}\}_{m=1}^M$. Once such information is conveyed to user d_l by BS s_l, this update can be performed at user d_l. After mobile user d_l updates $\mathbf{f}_l$, it sends them back to BS s_l. Next we analyze the step that updates $\mathbf{v}$ (cf. step (iii) in Section 3.3.2). In order to solve this problem locally at each BS s, the mobile users $d \in \left\{d_n \mid \ \forall\, l \in \mathrm{Out}(s) \bigcap \mathcal{L}^{wl}, n \in \bar{I}(l)\right\}$, i.e., the users whose transmissions are interfered by BS s, should send $\left(v_{nl} + \frac{\theta_{nl}}{\rho_2}\right)$ to BS s. After BS s obtains the updated v_l, it can broadcast these quantities back to those mobile users.

Given the above description of information exchange (summarized in Fig. 3.3), Algorithm 1 (and therefore, the N-MaxMin Algorithm) can be implemented in a distributed and parallel manner. The required information exchange can be reduced significantly if *a priori* knowledge about the paths used by the commodities is available.

For a C-RAN, there can be a few cloud centers, each responsible for updating the flow rates and precoders for a subnetwork of nodes. Suppose that the required channel state information is collected at the cloud centers, then the entire message passing can be made much more efficient. Specifically, only those variables belong to the links across different zones need to be exchanged. Within each subnetwork, a cloud center can execute its computational steps in parallel without any message exchange overhead. A more detail discussion on the implementation with subnetwork structure will be given in Chapter 4

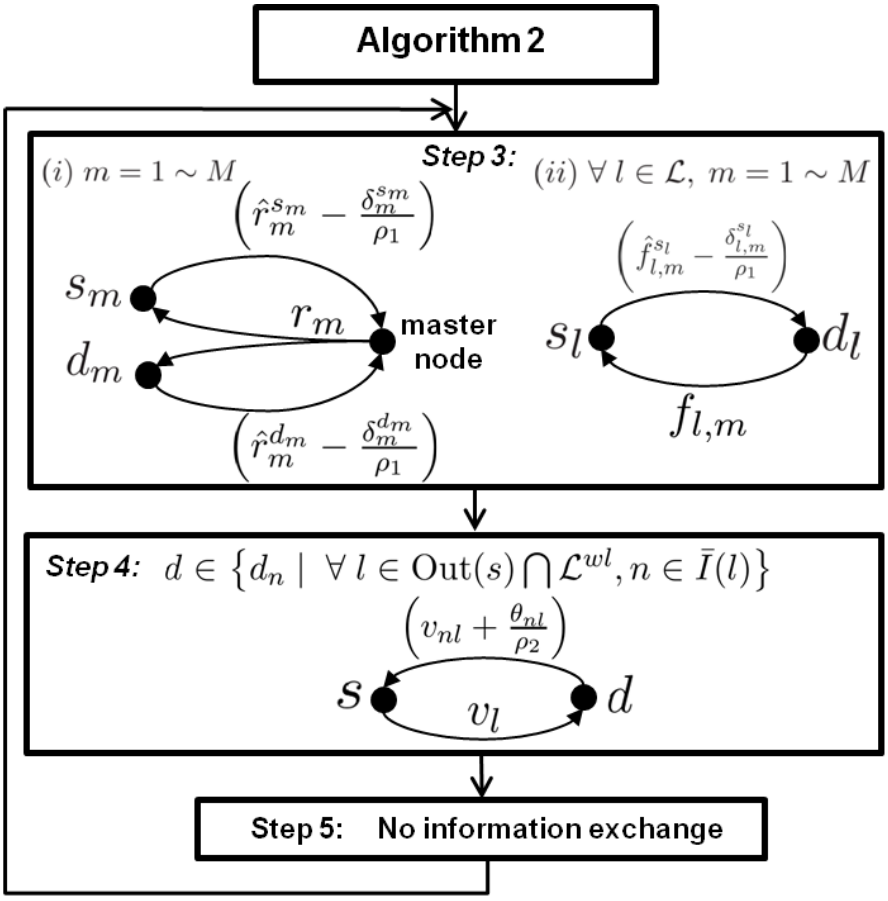


Figure 3.3: Summary of the required information exchange for each step of Algorithm 2.

3.3 Derivation of Updating Steps of Algorithm 1

In this section, we go over Algorithm 2 step by step and explain each of its update procedure. For notational simplicity, we ignore the superscript indices.

3.3.1 Solving Step 3 for Algorithm 2

In this step, problem (3.16) is solved to update $\{\mathbf{r}, \hat{\mathbf{v}}\}$. This problem can be further decomposed over the variables $\{r_{\min}, \{r_m\}_{m=1}^M, \mathbf{f}_l \mid \forall\, l \in \mathcal{L}^w\}$ and $\hat{\mathbf{v}} \cup \{\mathbf{f}_l \mid \forall\, l \in \mathcal{L}^{wl}\}$.

The first subblock only has to do with the wired links. A closer look at Step 3 of Algorithm 2 reveals that its optimization problem can be solved via two completely independent subproblems, one for variables $\{r_{\min}, \{r_m\}_{m=1}^M\}$ and the other for $\{\mathbf{f}_l \mid \forall\, l \in \mathcal{L}^w\}$. In the following we consider the two problems separately.

(i) Subproblem for $\{r_{\min}, \{r_m\}_{m=1}^M\}$: This step updates the current minimum flow rate among all commodities, and it can be mathematically expressed as

$$\begin{aligned} \max\ & \frac{r_{\min}}{2} - \frac{\rho_1}{2}\left(\hat{r}_{\min} - r_{\min} - \frac{\delta}{\rho_1}\right)^2 - \frac{\rho_1}{2}\sum_{m=1}^{M}\sum_{v\in\{s_m,d_m\}}\left(\hat{r}_m^v - r_m - \frac{\delta_m^v}{\rho_1}\right)^2 \\ \text{s.t. } & r_{\min} \geq 0,\ r_m \geq r_{\min},\ m = 1 \sim M. \end{aligned} \tag{3.19}$$

When $r_{\min}$ is fixed, the optimal $\{r_m^\star\}_{m=1}^M$ of problem (3.19) can be obtained by the first-order optimality condition, and this is expressed as

$$r_m^\star = \max\left\{r_m, \frac{1}{2}\sum_{v\in\{s_m,d_m\}}\left(\hat{r}_m^v - \frac{\delta_m^v}{\rho_1}\right)\right\}. \tag{3.20}$$

After plugging the obtained $r_m^\star$ back to the objective function of problem (3.19), the gradient of the objective function with respect to $r_{\min}$ is a decreasing function for $r_{\min} \geq 0$. The optimal $r_{\min}^\star = 0$ if the gradient is no more than 0 with $r_{\min} = 0$. Otherwise, it can be obtained such that the gradient equals 0.

(ii) Subproblem for $\{\mathbf{f}_l \mid \forall\, l \in \mathcal{L}^w\}$: It turns out that for this subset of variables, the corresponding updating procedure can be performed independently *over each link*.

For each link $l \in \mathcal{L}^w$, the following optimization problem is solved

$$\begin{aligned} \min \ & \sum_{m=1}^{M} \sum_{v \in \{s_l, d_l\}} \left(\hat{f}_{l,m}^v - f_{l,m} - \frac{\delta_{l,m}^v}{\rho_1} \right)^2 \\ \text{s.t. } & \mathbf{1}^T \mathbf{f}_l \leq C_l, \ \mathbf{f}_l \geq \mathbf{0}. \end{aligned} \tag{3.21}$$

The optimal solution $\mathbf{f}_l^\star$ can be obtained by the first-order optimality condition

$$f_{l,m}^\star = \frac{1}{2} \left[\sum_{v \in \{s_l, \, d_l\}} \left(\hat{f}_{l,m}^v - \frac{\delta_{l,m}^v}{\rho_1} \right) - \frac{\lambda_l^\star}{2} \right]^+, \tag{3.22}$$

where $\lambda_l^\star$ is the optimal Lagrange dual variable of the capacity constraint on link l. It can be obtained by bisection procedure such that the complementarity and feasibility condition for the capacity constraint are satisfied.

(iii) Subproblem for $\hat{\mathbf{v}} \cup \{\mathbf{f}_l \mid \forall \, l \in \mathcal{L}^{wl}\}$: The rest of variables are related only to the wireless links, and they are in fact decoupled across the wireless links. To be more specific, the problem for the wireless link $l = (s, d, f) \in \mathcal{L}^{wl}$ is shown below

$$\begin{aligned} \min \ & \frac{\rho_1}{2} \sum_{m=1}^{M} \sum_{v \in \{s_l, d_l\}} \left(\hat{f}_{l,m}^v - f_{l,m} - \frac{\delta_{l,m}^v}{\rho_1} \right)^2 + \frac{\rho_2}{2} \sum_{n \in I(l)} \left(v_n - v_{ln} - \frac{\theta_{ln}}{\rho_2} \right)^2 \\ \text{s.t. } & \mathbf{f}_l \geq \mathbf{0}, \text{ and (3.13)}. \end{aligned} \tag{3.23}$$

The optimal solution of this problem, $\{\mathbf{f}_l^\star, v_{ln}^\star \mid \ n \in I(l)\}$, can be obtained by the first-order conditions below

$$\begin{aligned} f_{l,m}^\star &= \frac{1}{2} \left[\sum_{v \in \{s_l, d_l\}} \left(\hat{f}_{l,m}^v - \frac{\delta_{l,m}^v}{\rho_1} \right) - \frac{\lambda_l^\star}{\rho_1} \right]^+, \\ v_{ln}^\star &= \begin{cases} \left(\rho_2 (v_l - \frac{\theta_{ll}}{\rho_2}) + \lambda_l^\star c_{2,l} \right) / \left(\rho_2 + 2\lambda_l^\star c_{3,ll} \right), & n = l \\ \rho_2 \left(v_n - \frac{\theta_{ln}}{\rho_2} \right) / \left(\rho_2 + 2\lambda_l^\star c_{3,ln} \right), & \forall \ n \in I(l) \setminus \{l\} \end{cases}. \end{aligned} \tag{3.24}$$

where $\lambda_l^\star$ is the optimal Lagrange dual variable for the rate-MSE constraint.

Plug the obtained optimal solutions (3.24) into the rate-MSE constraint of problem (3.23). The left hand side of the constraint is a decreasing function of $\lambda_l^\star$. Furthermore, the gradient of the right hand side of the rate-MSE constraint with respect to $\lambda_l^\star$ is nonnegative.

Hence, the right hand side of the rate-MSE constraint is an nondecreasing function of $\lambda_l^\star \geq 0$. Again, by the complementarity condition and the monotonicity of the rate-MSE constraint, the optimal value of $\lambda_l^\star$ can be computed via bisection search.

3.3.2 Solving Step 4 for Algorithm 2

The corresponding problem to update $\{\hat{\mathbf{r}}, \mathbf{v}\}$, i.e., step 4 of Algorithm 2, can be decomposed into two parts. The first part has to do with the flow rate conservation constraint with optimization variable $\hat{\mathbf{r}}$, and the second part has to do with $\mathbf{v}$.

The first part can again be separated into two independent subproblems, one for $\hat{r}_{\min}$ and another for the rest of the variables in $\hat{\mathbf{r}}$.

(i) Subproblem for $\hat{r}_{\min}$: The subproblem for variable $\hat{r}_{\min}$ is given by the following easy unconstraint quadratic optimization problem

$$\arg\max \ \frac{\hat{r}_{\min}}{2} - \frac{\rho_1}{2}\left(\hat{r}_{\min} - r_{\min} - \frac{\delta}{\rho_1}\right)^2 = r_{\min} + \frac{1+2\delta}{2\rho_1}. \tag{3.25}$$

(ii) Subproblem for $\{\hat{r}_m^{s_m}, \hat{r}_m^{d_m}, \hat{f}_{l,m}^{s_l}, \ \hat{f}_{l,m}^{d_l}\}$: In this subproblem, the rest of the variables in $\hat{\mathbf{r}}$ are updated, subject to the conservation constraints of flow rate. Since the introduction of the auxiliary local optimization variables, i.e., (3.10), this subproblem decoupled over each node $v \in \mathcal{V}$ and commodity m. As such, problem (3.17) decomposes into a series of simpler problems, one for each tuple (m, v)

$$\begin{aligned} \min \ & \sum_{l \in \mathrm{In}(v) \cup \mathrm{Out}(v)} \left(\hat{f}_{l,m}^v - f_{l,m} - \frac{\delta_{l,m}^v}{\rho_1}\right)^2 + 1_{\{s_m, d_m\}}(v)\left(\hat{r}_m^v - r_m - \frac{\delta_m^v}{\rho_1}\right)^2 \\ \text{s.t. } & (3.11). \end{aligned} \tag{3.26}$$

Since problem (3.26) has only one equality constraint, it admits a closed-form solution. In particular, denote the optimal dual Lagrangian variable as $\lambda_{v,m}^\star$. By the first-order optimality condition, the optimal solution for (3.26) is given by

$$\hat{f}_{l,m}^{v\star} = \begin{cases} f_{l,m} + \frac{\delta_{l,m}^v}{\rho_1} - \lambda_{v,m}^\star, & l \in \mathrm{Out}(v) \\ f_{l,m} + \frac{\delta_{l,m}^v}{\rho_1} + \lambda_{v,m}^\star, & l \in \mathrm{In}(v) \end{cases}, \tag{3.27a}$$

$$\hat{r}_m^{v\star} = \begin{cases} r_m + \frac{\delta_m^v}{\rho_1} - \lambda_{v,m}^\star, & v = d_m \\ r_m + \frac{\delta_m^v}{\rho_1} + \lambda_{v,m}^\star, & v = s_m \end{cases}, \tag{3.27b}$$

where $\lambda^{\star}_{v,m}$ is chosen to satisfy the flow conservation constraint.

(iii) Subproblem for v: The remaining part is for optimization variable **v** with power budget constraints, and this updating procedure can be decoupled over each BS. For BS $s \in \mathcal{B}$, the updating rule is,

$$\begin{aligned} \min \quad & \sum_{l \in \text{ Out}(s) \bigcap \mathcal{L}^{wl},\, n \in \bar{I}(l)} \left(v_l - v_{nl} - \frac{\theta_{nl}}{\rho_2} \right)^2 \\ \text{s.t.} \quad & \sum_{l \in \text{ Out}(s) \bigcap \mathcal{L}^{wl}} v_l^2 \leq P_s. \end{aligned} \tag{3.28}$$

By denoting the optimal Lagrange dual variable for the power constraint as $\lambda^{\star}_s \geq 0$ and the optimal solution of problem (3.28) as $\{v^{\star}_l \mid l \in \mathcal{L}^{wl}\}$, the optimal solutions can be expressed as

$$v^{\star}_l = \frac{\sum_{n \in \bar{I}(l)} v_{nl} + \frac{\theta_{nl}}{\rho_2}}{|\bar{I}(l)| + \lambda^{\star}_s}. \tag{3.29}$$

One can observe that $\sum_{l \in \text{ Out}(s) \bigcap \mathcal{L}^{wl}} |v^{\star}_l|^2$ is a decreasing function of $\lambda^{\star}_s$. So, $\lambda^{\star}_s$ can be chosen via a bisection search to ensure the complementarity and feasibility conditions for the power budget constraint.

To summarize, all the steps in Algorithm 2 (including the updating of the Lagrange dual variables, (3.18)) can be efficiently computed.

3.4 Numerical Experiments

In this section, we report some numerical results on the performance of the proposed algorithms as applied to a network with 57 BSs and 11 network routers. We have tested both the the efficacy and the efficiency of the proposed algorithms. The topology and the connectivity of this network are shown in Fig. 4.7. For the backhaul links of this network, a fixed capacity is assumed, and is same in both directions. These link capacities are given as follows:

- links between routers and those between gateway BSs and the routers: 1 (Gnats/s);
- 1-hop to the gateways: 100 (Mnats/s);
- 2-hop to the gateways: [10,50] (Mnats/s);

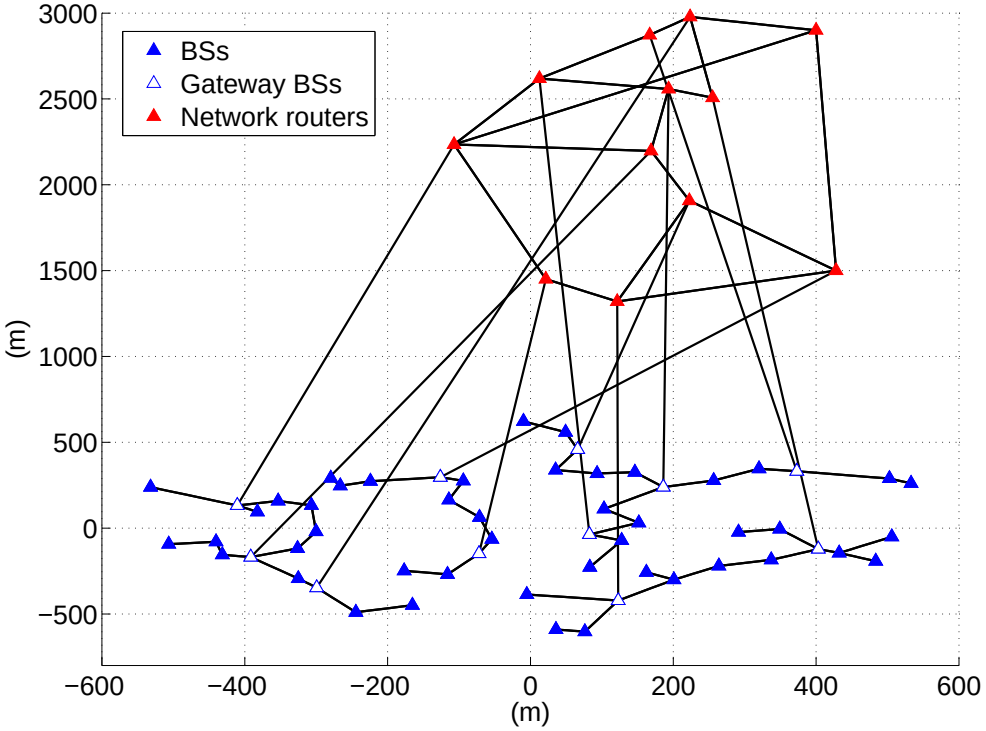


Figure 3.4: The considered network consists of 57 BSs and 11 routers with the locations and the connectivity between these nodes.

- 3-hop to the gateways: [2,5] (Mnats/s);
- More than 4-hop to the gateways: 0 (nats/s).

The number of subchannels is $F = 3$ and each subchannel has 1 MHz bandwidth. The power budget for each BS is chosen equally by $P = P_s$, $\forall\ s \in \mathcal{B}$, and $\sigma_d^2 = 1$, $\forall\ d \in \mathcal{U}$. The wireless links follow the Rayleigh distribution with $CN(0, (200/\text{dist})^3)$, where dist is the distance between BS and the corresponding user. The source (destination) node of each commodity is randomly selected from network routers (mobile users), and all simulation results are averaged over 100 randomly selected end-to-end commodity pairs. Below we refer to one round of the N-MaxMin iteration as an *outer iteration*, and one round of Algorithm 2 for solving $(\mathbf{r}, \mathbf{v})$ as an *inner iteration*.

In the first experiment, we assume that all mobile users can be served by BSs within 300 meters and are interfered by all the BSs. For this problem, the parameters of N-MaxMin algorithm are set to be $\rho_1 = 0.1$ and $\rho_2 = 0.1$, 0.05, and $0,01$ for, respectively, $p = 0$dB, 10dB, and 20dB; the termination criterion is

$$\frac{(r_{\min}^{(t+1)} + \hat{r}_{\min}^{(t+1)}) - (r_{\min}^{(t)} + \hat{r}_{\min}^{(t)})}{r_{\min}^{(t)} + \hat{r}_{\min}^{(t)}} < 10^{-3}$$
$$\max\{\|\mathbf{C}\mathbf{r}^{(t)} - \hat{\mathbf{r}}^{(t)}\|_\infty, \|(\mathbf{D}\mathbf{v}^{(t)})^2 - (\hat{\mathbf{v}}^{(t)})^2\|_\infty\} < 5 \times 10^{-4}$$

where $(\cdot)^2$ represents elementwise square operation.

For the comparison purpose, the following two heuristic algorithms are considered.

Heuristic 1 (greedy approach):

We assume that each mobile user is served by a single BS on a specific frequency tone. For each user, we pick the BS and channel pair that has the strongest channel as its serving BS and channel. After BS-user association is determined, each BS uniformly allocates its power budget to the available frequency tones as well as to the served users on each tone.

With the obtained power allocation and BS-user association, the capacity of all wireless links are available and fixed, so the min-rate of all commodities can be maximized by solving a wireline routing problem.

Heuristic 2 (orthogonal wireless transmission):

For the second heuristic algorithm, each BS uniformly allocates its power budget to each frequency tone. To obtain a tractable problem formulation, we further assume that each active wireless link is interference free. Hence, each wireless link rate constraints now becomes convex. To impose this interference free constraint, additional variables $\beta_l \in \{0, 1\},\ \forall\, l \in \mathcal{L}^{wl}$ are introduced, where $\beta_l = 1$ if wireless link l is active, otherwise $\beta_l = 0$. In this way, there is no interference on wireless link l if $\sum_{n \in I(l)} \beta_n = 1$. To

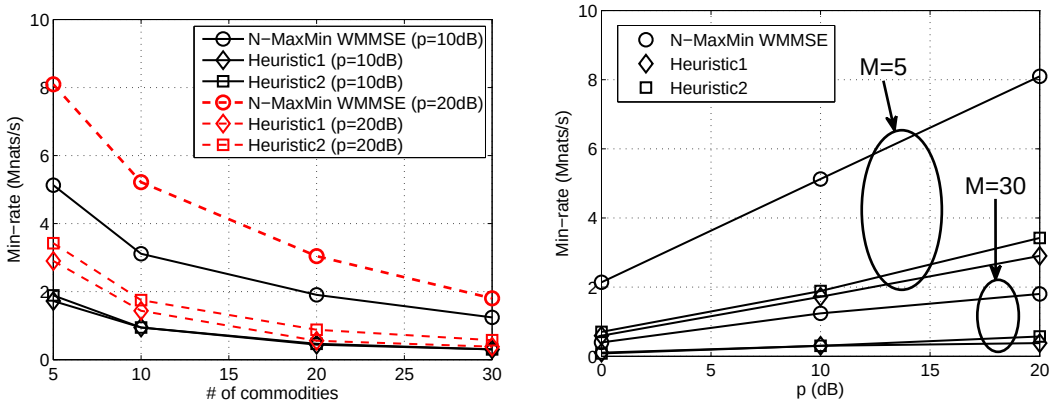


Figure 3.5: The min-rate achieved by N-MaxMin algorithm and the two heuristic algorithms for different M and power budgets.

summarize, we solve the following optimization problem:

$$\begin{aligned} &\max \ r_{\min} \\ &\text{s.t. } \mathbf{1}^T \mathbf{f}_l \leq \beta_l \log\left(1 + \frac{|h_l|^2 \bar{p}_{s_l}/K}{\sigma_{d_l}^2}\right), \\ &\qquad \sum_{n \in I(l)} \beta_n = 1, \ \beta_l \in \{0,1\}, \ \forall\, l, n \in \mathcal{L}^{wl}, \\ &\qquad (3.1), \ (3.3), \text{ and } (3.5b). \end{aligned}$$

Since the integer constraints on $\{\beta_l \mid \forall\ l \in \mathcal{L}^{wl}\}$ are also intractable, we relax it to $\beta_l = [0,1]$. In this way the problem becomes a large-scale LP, whose solution represents an upper bound value of this heuristic.

In Fig. 3.5, we show the min-rate performance of different algorithms for different number of commodities and power budget. We observe that the min-rate achieved by the N-MaxMin algorithm is more than twice of those achieved by the heuristic algorithms.

In the second set of numerical experiments, we evaluate the proposed N-MaxMin algorithm using different number of commodity pairs and different power budgets at the BSs. Here we use the same settings as in the previous experiment, except that all mobile users are interfered by the BSs within a distance of 800 meters, and that we

set $\rho_2 = 0.005$ (resp. $\rho_2 = 0.001$) when $P = 10$ dB (resp. $P = 20$ dB). The min-rate performance for the N-MaxMin algorithm and the required number of inner iterations are plotted in Fig. 3.6. Due to the fact that the obtained $\{\mathbf{r}, \mathbf{v}\}$ is far from the stationary solution in the first few outer iterations, there is no need to complete Algorithm 2 at the very beginning. Hence, we limit the number of inner iterations to be no more than 500 for the first 5 outer iterations. After the early termination of the inner Algorithm 2, we use the obtained $\mathbf{v}$ to update $\mathbf{u}$ and $\mathbf{w}$ by (3.8) and (3.9), respectively.

In Fig. 3.6(a)–(b), we see that when $P = 10$ dB, the min-rate converges at about the 10th outer iteration when the number of commodities is up to 30, while less than 500 inner iterations are needed per outer iteration. Moreover, after the 10th outer iteration, the number of inner ADMM iterations reaches below 100. In Fig. 3.6(c)–(d), the case with $P = 20$dB is considered. Clearly the required number of outer iterations is slightly more than that in the case of $P = 10$dB, since the objective value and the feasible set are both larger. However, in all cases the algorithm still converges fairly quickly. Also, for a cloud-based C-RAN architecture, the network nodes are partitioned into several subnetworks, each managed by a separate cloud center. In this case, the computation can be distributed across the cloud centers, with the communication overhead restricted to only the variables associated with the links connecting the neighboring subnetworks.

In the last set of numerical experiments, we demonstrate how the parallel implementation can speed up Algorithm 2 considerably. To illustrate the benefit of parallelization, we consider a larger network (see Fig. 3.7) which is derived by merging two identical BS networks shown in Fig. 4.7. The new network consists of 126 nodes (12 network routers and 114 BSs). For simplicity, we removed all the wireless links, so constraints (3.2) and (3.4) of problem (3.5) are absent. This reduces problem (3.5) to a network flow problem (a very large linear program).

We implement Algorithm 2 using the Open MPI package, and compare its efficiency with the commercial LP solver, Gurobi [121]. For the Open MPI implementation, we use 9 computation cores for each set of network nodes as illustrated in Fig. 3.7. We choose $\rho_1 = 0.01$ and let the BSs serve as the destination nodes for commodities. Table 4.3 compares the computation time required for different implementation of Algorithm 2 and that of Gurobi. We observe that parallel implementation of Algorithm 2 leads to more than 5 fold improvement in computation time computed on SunFire X4600

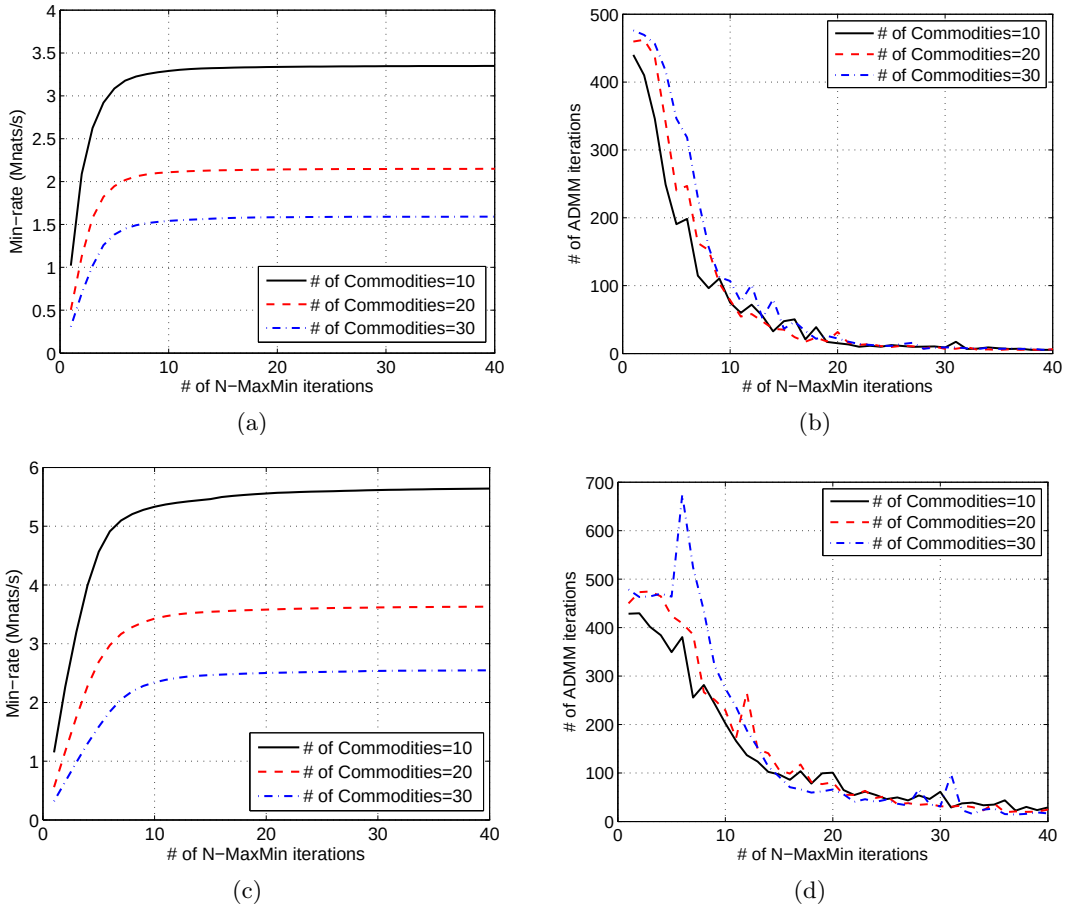


Figure 3.6: The min-rate performance and the required number of iterations for the proposed N-MaxMin algorithm. In [(a)(b)] $P = 10$dB and in [(c)(d)] $P = 20$dB. In [(a)(c)], the obtained min-rate versus the iterations of N-MaxMin is plotted. In [(b)(d)], the required number of inner ADMM iterations is plotted against the iteration for the outer N-MaxMin algorithm.

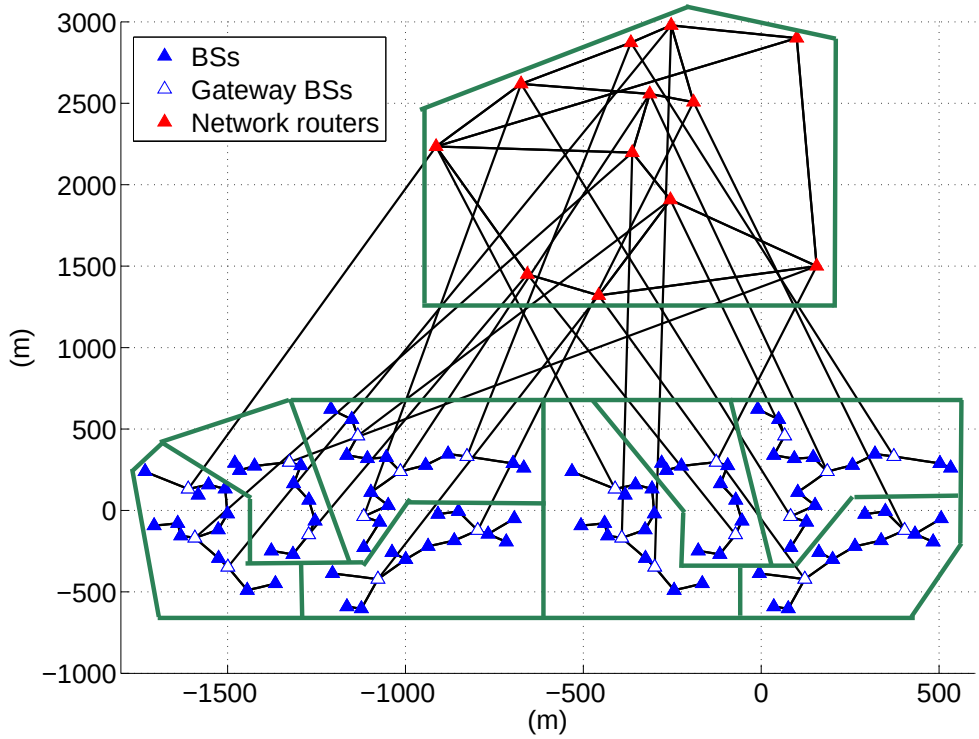


Figure 3.7: The considered network consists of 114 BSs and 11 routers with the locations and the connectivity between these nodes. Each computation core is responsible for one group of nodes shown in the figure

server with AMD Opteron 8356 2.3GHz CPUs. We also note that when the problem size increases, the performance of Gurobi becomes worse than that achieved by the parallel implementation of Algorithm 2. Thus, the proposed algorithm (implemented in parallel) appears to scale nicely to large problem sizes.

In the last numerical experiment, we have demonstrated the computational benefits for the proposed ADMM-based algorithm via distributed and parallel implementation. However, the proposed distributed algorithm 1 (c.f. Chapter 2) and 2 (c.f. Chapter 3) demand each computation node to be synchronous to each other. In the next chapter, we will further relax this synchronization requirement, which can further speed-up the proposed distributed algorithm.

# of Commodities	50	100	300
Time (s) (Sequential)	1.04	2.03	8.53
Time (s) (Parallel)	0.20	0.37	1.10
Time (s) (Gurobi)	0.20	0.64	2.51
# of Variables	1.4×10^4	2.9×10^4	8.7×10^4
# of Constraints	2.1×10^4	4.2×10^4	1.3×10^5

Table 3.3: Comparison of computation time used by different implementations for the routing only problem.

Chapter 4

A Distributed Semi-Asynchronous Algorithm for the provision of C-RAN

The ADMM has been popular for solving large-scale convex optimization problems. Among all its features, ADMM is easily implementable over a network of distributed nodes, making it the state-of-the-art algorithm for large-scale distributed optimization. For instance, in the previous chapters, we have proposed Algorithm 1 and 2 for the distributed implementation. However, these algorithms require the synchronization between the computation nodes. In this chapter, we propose an asynchronous distributed algorithm that can handle a specific form of asynchronism arising in the network. Specifically, our proposed algorithm is based on the so-called BSUM-M algorithm [108], a new variant of ADMM. Theoretically, we show that the proposed algorithm converges to the global optimal solution under some assumptions on the degree of network asynchrony. Practically, the effectiveness and efficiency of the proposed asynchronous algorithm are illustrated through solving the backhaul network routing problem of the C-RAN architecture.

4.1 Related Works

Asynchronous algorithms such as those studied in the classical work of [38] are less susceptible to network synchronization error and are more robust to communication failures. Hence they are of great interests to distributed big data processing. Recently, many different asynchronous first-order algorithms have been proposed [122–125], under different assumptions on the network asynchrony. Specifically, [122] allows the gradients to be updated with random subset of nodes, thus the overall performance of the system will not be much affected by temporary node failures. In [123–125], the communication delays between computation nodes have been taken into account, allowing the gradient at each computation node to be calculated with (possibly) outdated information from other computation nodes. These algorithms [123–125] have been shown to converge even in the dynamic scenario.

A different paradigmn for distributed optimization is the ADMM, which has found applications in a plethora of machine learning and networking problems. Many variants of ADMM have been proposed in the literature [108, 126–128]. However, only a few asynchronous versions of ADMM have been proposed [109–112, 129]. In [111], a randomized version of ADMM is considered where a random subset of nodes are updated at each iteration, a form of network asynchrony closely related to that considered in [122]. In [112], another randomized ADMM is considered, but it further incorporates the outdatedness of information between computation nodes. In [109, 110], both the heterogeneity of computation nodes and communication delays/failures have been incorporated into the ADMM framework for the so-called global consensus problem [39]. Furthermore, the algorithm in [110] can deal with not only convex problems but also a class of nonconvex ones. Although these two algorithms accommodate network asynchrony, they are only shown to work for special optimization problems with consensus constraint among nodes. Moreover, for [109], no constraint is allowed for the consensus problem, and for [110], all nonsmooth terms and constraints should be handled by the central computation node.

In this chapter, we propose one specific asynchronous distributed algorithm based on the BSUM-M recently developed in [108], which is a variant of ADMM that handles multiple blocks of variables. Our approach follows the *semi-asynchronous* scheme

in which the computation nodes update a subset of its variables that are coupled to the new incoming information from other nodes. The semi-asynchronous scheme differs from the well-known *partially asynchronous* scheme described in [38] in the sense that no out-of-sequence communication is allowed. Moreover, the scheme is very similar to the asynchronous implementation in [111, 112]. However, in [111, 112], the sequence of variable updates follows a random distribution, which cannot well model the asynchronism due to different processing speed between computation nodes. In contrast, the semi-asynchronous scheme studied in this chapter can be viewed as a more practical deterministic counterpart, in which the nodes are updated following an *essentially cyclic* (EC) rule [91, 113]. Furthermore, unlike the existing asynchronous ADMM algorithms [109,110], the proposed asynchronous algorithm allows each computation node to have its own local constraints. The latter feature gives extra design flexibility to distribute the computation loads across different local nodes.

4.2 System Model and Problem Formulation

In this chapter, the following structured convex optimization problem is considered:

$$\min_{\boldsymbol{x}} \quad f(\boldsymbol{x}) = \sum_{i=0}^{K} f_i(\boldsymbol{x}_i) \tag{4.1a}$$

$$\text{s.t.} \quad \mathbf{A}_{ij}^{i}\boldsymbol{x}_i + \mathbf{A}_{ij}^{j}\boldsymbol{x}_j = \mathbf{b}_{ij},\ 0 \le i < j \le K, \tag{4.1b}$$

$$\boldsymbol{x}_i \in \mathcal{X}_i,\ i = 0 \sim K, \tag{4.1c}$$

where the optimization variable $\boldsymbol{x} \in \mathbb{R}^n$ is partitioned as $\boldsymbol{x} = \left[\boldsymbol{x}_0^T, \ldots, \boldsymbol{x}_K^T\right]^T$ with $\boldsymbol{x}_i \in \mathbb{R}^{n_i}$, $i = 0 \sim K$, and $\sum_{i=0}^{K} n_i = n$; $\mathcal{X}_i$, $i = 0 \sim K$, is the convex feasible set for $\boldsymbol{x}_i$ and $\mathcal{X} = \prod_{i=0}^{K} \mathcal{X}_i$; $\mathbf{A}_{ij}^{i} \in \mathbb{R}^{m_{ij} \times n_i}$ and $\mathbf{A}_{ij}^{j} \in \mathbb{R}^{m_{ij} \times n_j}$, $0 \le i < j \le K$, are arbitrary matrices that couple variables $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$; f_i, $i = 0 \sim K$, is a smooth convex function. Many contemporary problems, especially problems with underlying network structures, in signal processing, machine learning and communication systems can be formulated in the form of (4.1); see e.g., [39]. To solve problem (4.1), we start with the augmented

Lagrangian function of (4.1) as follows:

$$L(\boldsymbol{x};\boldsymbol{\lambda}) = \sum_{i=0}^{K} f_i(\boldsymbol{x}_i) + \sum_{0\le i<j\le K} \Big[\langle \mathbf{A}_{ij}^i \boldsymbol{x}_i + \mathbf{A}_{ij}^j \boldsymbol{x}_j - \mathbf{b}_{ij}, \boldsymbol{\lambda}_{ij}\rangle + \frac{\rho}{2}\|\mathbf{A}_{ij}^i \boldsymbol{x}_i + \mathbf{A}_{ij}^j \boldsymbol{x}_j - \mathbf{b}_{ij}\|^2\Big] \tag{4.2}$$

where $\rho > 0$ is the augmented Lagrangian parameter; $\boldsymbol{\lambda}_{ij} \in \mathbb{R}^{m_{ij}}$, $0 \le i < j \le K$, is the Lagrangian dual variables for linear equality constraint (4.1b), and we will denote $\boldsymbol{\lambda} \triangleq [\{\boldsymbol{\lambda}_{ij}^T\}_{ij}]^T$. The ADMM algorithm then solves problem (4.1) by

$$\boldsymbol{x}_i^{t+1} = \arg\min_{\boldsymbol{x}_i \in \mathcal{X}_i} L(\boldsymbol{x}_0^{t+1}, \ldots, \boldsymbol{x}_{i-1}^{t+1}, \boldsymbol{x}_i, \boldsymbol{x}_{i+1}^{t}, \ldots, \boldsymbol{x}_K^{t}; \boldsymbol{\lambda}^t),\ i = 0 \sim K, \tag{4.3a}$$

$$\boldsymbol{\lambda}_{ij}^{t+1} = \boldsymbol{\lambda}_{ij}^{t} + \rho(\mathbf{A}_{ij}^i \boldsymbol{x}_i^{t+1} + \mathbf{A}_{ij}^j \boldsymbol{x}_j^{t+1} - \mathbf{b}_{ij}),\ 0 \le i < j \le K, \tag{4.3b}$$

where $t = 1, 2, 3\ldots$ is the iteration index. For the special case that $K = 1$, the update procedure (4.3) reduces to the 2-block ADMM algorithm, which can converge to the optimal primal and dual solutions, i.e., $\boldsymbol{x}^\star$ and $\boldsymbol{\lambda}^\star$, under some mild conditions [38, Proposition 4.2]. Here $\boldsymbol{\lambda}^\star$ solves the dual problem $\max_{\boldsymbol{\lambda}}[\min_{\boldsymbol{x}\in\mathcal{X}} L(\boldsymbol{x};\boldsymbol{\lambda})]$, which achieves the same optimal dual value as the global minimum of problem (4.1) if strong duality holds [33]. However, when $K > 1$, the multi-block ADMM has been shown to not always converge [130], and some variants have been proposed to guarantee convergence, e.g., [108, 128, 131].

In the following, we make some standard assumptions for the considered problem (4.1),

(A1) The global minimum and dual optimal value of problem (4.1) is attainable. The intersection $\mathcal{X} \cap \mathrm{int}(\mathrm{dom}\ f) \cap \{\boldsymbol{x} \mid (4.1\mathrm{b})\}$ is nonempty.

(A2) $f_i(\boldsymbol{x}_i)$, $i = 0 \sim K$, has Lipchitz continuous gradient, i.e.,

$$\|\nabla f_i(\mathbf{v}) - \nabla f_i(\mathbf{u})\| \le L\|\mathbf{v} - \mathbf{u}\|,\ \forall\ \mathbf{v}, \mathbf{u} \in \mathcal{X}_i. \tag{4.4}$$

(A3) The function $f_i(\boldsymbol{x}_i)$, $i = 0 \sim K$, can be decomposed as $f_i(\boldsymbol{x}_i) = g_i(\mathbf{D}_i\boldsymbol{x}_i) + \langle \boldsymbol{x}_i, \mathbf{b}\rangle$, where $\mathbf{D}_i$ is some given matrix (not necessarily full column rank), and $g_i(\cdot)$ is a strongly convex and continuously differentiable function on $\mathrm{int}(\mathrm{dom}\ g_i)$, i.e., there exists some $\delta_i > 0$ some that

$$g_i(\mathbf{v}) - g_i(\mathbf{u}) \ge \langle \nabla g_i(\mathbf{u}), \mathbf{v} - \mathbf{u}\rangle + \frac{\delta_i}{2}\|\mathbf{v} - \mathbf{u}\|^2, \quad \forall\ \mathbf{v}, \mathbf{u} \in \mathrm{int}(\mathrm{dom}\ g_i).$$

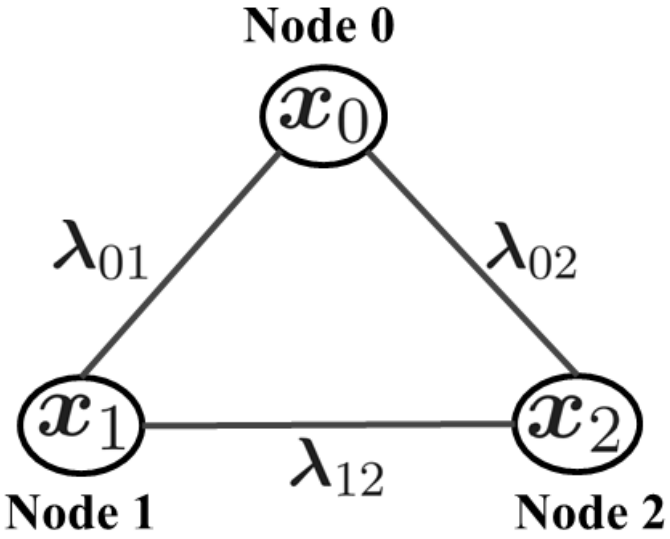


Figure 4.1: An illustrative diagram of the considered distributed system.

(A4) The feasible sets $\mathcal{X}_i$, $i = 0 \sim K$, are compact polyhedral sets, and are given by $\mathcal{X}_i \triangleq \{\boldsymbol{x}_i \mid \mathbf{C}_i\boldsymbol{x}_i \leq c_i\}$, for some matrix $\mathbf{C}_i$ and c_i.

Note that, the assumptions (A3) and (A4) can be replaced by (A3'): $f_i(\cdot)$ is strongly convex, $i = 0 \sim K$.

Given the problem formulation (4.1) and the above assumptions, our goal is to solve it distributedly with some tolerance for asynchronism. Specifically, we consider a network topology where there are $K+1$ computation nodes. We let each node i, $i = 0, 1, \cdots, K$, keeps $\boldsymbol{x}_i$. Every pair of node i and j, $0 \leq i < j \leq K$, exchanges $\mathbf{A}_{ij}^i\boldsymbol{x}_i$ and $\mathbf{A}_{ij}^j\boldsymbol{x}_j$ while the Lagrangian dual variable $\boldsymbol{\lambda}_{ij}$ is shared between them for solving problem (4.1). See Fig. 4.1 for the illustration of the considered distributed system with 3 computation nodes. Most of the traditional distributed implementation, e.g., the ADMM approach in (4.3), requires the synchronization between the nodes, e.g., each node cannot perform any update until it receives the *latest* information from *all* other neighboring nodes. This means the efficiency of the synchronous implementation strongly depends on the communication delays, and if the nodes have different computational capabilities it will be dominated by the slowest computing node. In Fig. 4.2(a), an example of the synchronous implementation for the distributed system in Fig. 4.1 is provided with node 2 being the slowest node.

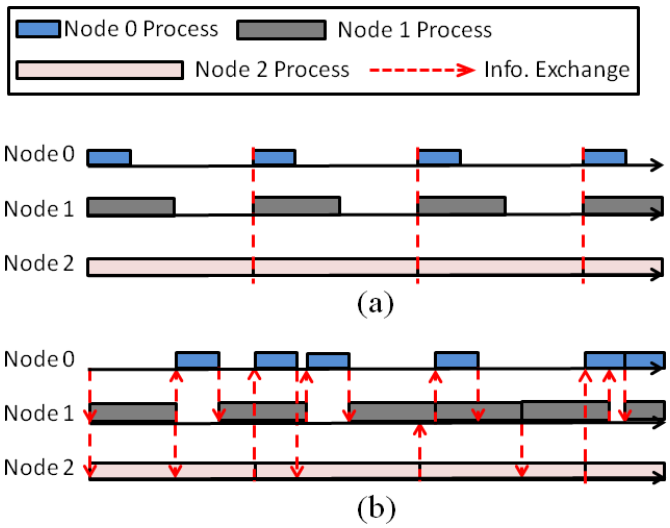


Figure 4.2: The illustration of different updating schemes with 3 nodes. (a) synchronous; (b) proposed semi-asynchronous.

In this chapter, we consider the semi-asynchronous scheme which allows the use of *outdated* information at the nodes. Specifically, as illustrated in Fig. 4.2(b), each node starts to update its own local variable whenever an updated information from a *subset* of other nodes is received. For example, in Fig. 4.2(b), the second update of node 1 occurs when it receives the information from node 0. However, during its update, the information of $\mathbf{A}_{01}^{0}\boldsymbol{x}_0$ and $\mathbf{A}_{12}^{2}\boldsymbol{x}_2$ in node 1 are becoming outdated since $\boldsymbol{x}_0$ and $\boldsymbol{x}_2$, are, respectively, being updated at the same time in node 0 and node 2. As is illustrated in the figure, we assume that the order of the communication between nodes is not out-of-sequence in the proposed semi-asynchronous scheme. This model is more restrictive than the partially asynchronous scheme defined in [38]. Since no synchronization requirement is needed for the proposed semi-asynchronous scheme, the computation capabiilty of each node would not be wasted. The efficiency of the semi-asynchronous approach would hence be improved over the synchronous counterpart, especially when the computational speeds are not balanced across the nodes. Before going into the details of the proposed approach, we discuss in the next section a motivating application on the design of TE algorithm for a hierarchical network.

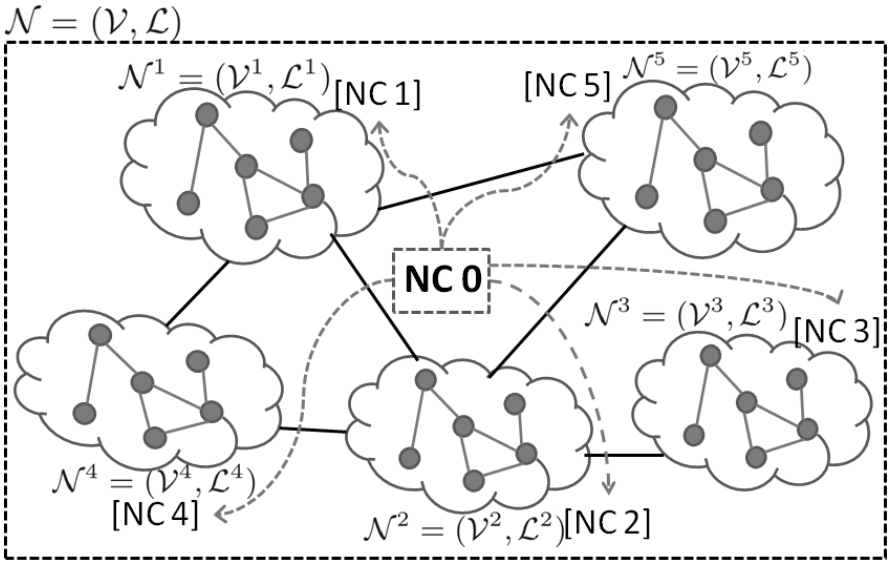


Figure 4.3: A wireline network consists of 5 subnetworks. Each of them is controlled by a network controller (NC), and these NCs are coordinated globally by a central NC 0.

4.3 An Application: Hierarchical Network Traffic Engineering

As an application, we apply the proposed semi-asynchronous BSUM-M algorithm to the hierarchical network traffic engineering (TE) problem, in which a number of network controllers (NCs) are deployed in different geographical locations, each controlling a set of network nodes. A master NC, say NC 0, globally coordinates the behavior of the distributed NCs, see Fig. 4.3 for an illustration. Such network appears for example in the backhaul of the C-RAN architecture in Chapter 3, in which each NC is a cloud center managing a subset of closely located network routers and base stations.

We consider a connected network $\mathcal{N} = (\mathcal{V}, \mathcal{L})$ which is controlled by $K+1$ NCs. Let $\mathcal{V}$ denote the set of network nodes, which is partitioned into K subsets, i.e., $\mathcal{V} = \bigcup_{i=1}^{K} \mathcal{V}^i$, $\mathcal{V}^i \cap \mathcal{V}^j = \emptyset$, $\forall\ i \neq j$. The set of directed links is denoted as $\mathcal{L} \triangleq \{l = (s_l, d_l) \mid \forall\ s_l, d_l \in \mathcal{V}\}$, where $l = (s_l, d_l)$ denotes the directed link from node s_l to node d_l. Each NC i, $i = 1 \sim K$, controls the ith subnetwork $\mathcal{N}^i$ which consists of $\mathcal{V}^i$ and the links connecting these nodes, i.e., $\mathcal{L}^i \triangleq \{l = (s_l, d_l) \in \mathcal{L} \mid \forall\ s_l, d_l \in \mathcal{V}^i\}$. Also define $\mathcal{L}^0_{ij} \triangleq \{l = (s_l, d_l) \in \mathcal{L} \mid \forall\ s_l \in \mathcal{V}^i, d_l \in \mathcal{V}^j\}$ as the set of links connecting two

neighboring subnetworks i and j. The links connecting different subnetworks is denoted as $\mathcal{L}^0 = \bigcup_{i \neq j} \mathcal{L}^0_{ij}$. We also assume a master node, denoted by NC 0, exists which controls the fairness between all the data flow rates $\{r_m\}_{m=1}^M$.

Consider maximizing the minimum rate of all data flows. Following the same notations as in Chapter 3, the problem can be formulated as the following linear program

$$\min_{\mathbf{f},\, \mathbf{r}} \quad -r_{\min} \quad \text{s.t.} \quad \mathbf{f} \geq 0,\ r_m \geq r_{\min},\ m = 1 \sim M \tag{4.5a}$$

$$(3.1) \text{ and } (3.3), \tag{4.5b}$$

where constraints (3.1) and (3.3) are, respectively, for the link capacity and per-node flow conservation condition, see Chapter 3.1 for details; and $\mathbf{f} \triangleq \{\mathbf{f}_l \mid l \in \mathcal{L}\}$ and $\mathbf{r} \triangleq \{r_{\min}, r_m \mid m = 1 \sim M\}$. Note that the minimum rate is used here because it assures a fair rate allocation among the data flows, and such utility has been adopted by many recent works; e.g., [119]. Other objective functions such as the sum rate or the proportional fairness can be used here as well.

To put problem (4.5) into the form of (4.1), we introduce a few auxiliary variables (see Fig. 4.4 for illustration)

- For each flow rate $\mathbf{f}_l \in \mathcal{L}^0_{ij}$, $\forall\ i, j$, we introduce *two* extra copies, namely $\mathbf{f}_l^{s_l}$ and $\mathbf{f}_l^{d_l}$. We let $\mathbf{f}_l$ be controlled by NC i, and $\mathbf{f}_l^{s_l}$ and $\mathbf{f}_l^{d_l}$ be individually managed by the two neighboring NC i and j.

- For each data flow rate r_m, we introduce *two* extra copies: $r_m^{s_m}$ and $r_m^{d_m}$. The original one r_m is managed by the master node while $r_m^{s_m}$ and $r_m^{d_m}$ are, respectively, managed by the source and the destination NCs of flow m.

- Within each subnetwork, we introduce a new copy $\tilde{\mathbf{f}}_l$ for the link flow rate $\mathbf{f}_l$, $\forall\ l \in \mathcal{L} \setminus \mathcal{L}^0$.

For notational simplicity, we have created a few groups of the variables and denote them as $\{\boldsymbol{x}_i\}_i$, $\{\boldsymbol{x}_{0i}\}_i$, $\{\boldsymbol{x}_{i1j}\}_{i,j\neq i}$, $\{\boldsymbol{x}_{i2j}\}_{ij}$,$\{\boldsymbol{x}_{i3}\}_i$, and $\{\boldsymbol{x}_{i4}\}_i$; see Table 4.1 for detailed definitions.

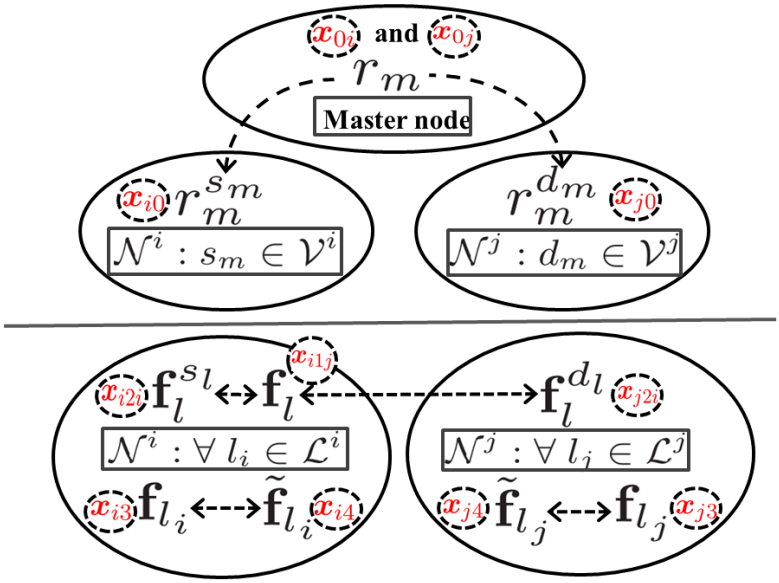


Figure 4.4: The relationship across the introduced local auxiliary variables. The variables connected by dash lines should be equal to each other. Each variable belongs to the variable group inside the closest dash circle.

Obviously, the original variables and their splits should be identical, therefore we have the following sets of equality constraints

$$\underbrace{\boldsymbol{x}_{0i} = \boldsymbol{x}_{i0} \Rightarrow \mathbf{A}^0_{0i}\boldsymbol{x}_0 = \mathbf{A}^i_{i0}\boldsymbol{x}_i}_{\text{in master node and } \mathcal{N}^i},\ i = 1 \sim K, \tag{4.6a}$$

$$\underbrace{\boldsymbol{x}_{i1j} = \boldsymbol{x}_{j2i} \text{ and } \boldsymbol{x}_{j1i} = \boldsymbol{x}_{i2j} \Rightarrow \mathbf{A}^i_{ij}\boldsymbol{x}_i = \mathbf{A}^j_{ij}\boldsymbol{x}_j}_{\text{in } \mathcal{N}^i \text{ and } \mathcal{N}^j},\ 1 \le i < j \le K, \tag{4.6b}$$

$$\underbrace{\boldsymbol{x}_{i2i} = \{\boldsymbol{x}_{i1j}\}_{j\neq i},\ \boldsymbol{x}_{i3} = \boldsymbol{x}_{i4}}_{\text{in } \mathcal{N}^i},\ i = 1 \sim K, \tag{4.6c}$$

where $\mathbf{A}_i$ and $\mathbf{A}_{ij}$, $0 \le i < j \le K$, are the matrices that ensure the compact expressions for the equality constraints are the same as the original ones. For example, in (4.6a), each row of $\mathbf{A}_{0i}$ and $\mathbf{A}_{i0}$ is a vector, which has all elements equal to 0 except for one element which equals 1. Thus, each component of $\mathbf{A}_{0i}\boldsymbol{x}_0$ is a specific element of $\boldsymbol{x}_{0i}$, which is a subvector of $\boldsymbol{x}_0$. Similarly argument applies for $\mathbf{A}_{i0}\boldsymbol{x}_i$. By properly allocating the variables $\boldsymbol{x} = \{\boldsymbol{x}_i\}_{i=0\sim K}$ to the constraints of problem (4.5), these constraints become separable over subnetworks. Moreover, (3.1) and (3.3) become independent to each

Notations	Definitions	Physical meaning
$\boldsymbol{x}_0$	$\{r_{\min}\} \cup \left(\cup_{i=1}^{K} \boldsymbol{x}_{0i}\right)$	The variables stored in the master node
$\boldsymbol{x}_{0i}$	$\{r_m \mid \forall\, m \text{ s.t. } s_m \text{ or } d_m \in \mathcal{V}^i\}$	The data flow rates originate from or go to $\mathcal{N}^i$
$\boldsymbol{x}_i$	$\boldsymbol{x}_{i0} \cup \{\boldsymbol{x}_{i1j}\}_{j\neq i} \cup \{\boldsymbol{x}_{i2j}\}_j \cup \boldsymbol{x}_{i3} \cup \boldsymbol{x}_{i4}$	The variables stored in $\mathcal{N}^i$
$\boldsymbol{x}_{i0}$	$\{r_m^v \mid v \in \mathcal{V}^i, \forall\, m\}$	The auxiliary variables copied from $\boldsymbol{x}_{0i}$ in the master node
$\boldsymbol{x}_{i1j},\ j \neq i$	$\{\mathbf{f}_l \mid l \in \mathcal{L}_{ij}^0\}$	The bordering flow rate variables for links $l \in \mathcal{L}_{ij}^0$
$\boldsymbol{x}_{i2j},\ j \neq i$	$\{\mathbf{f}_l^v \mid l \in \mathcal{L}_{ji}^0, v \in \mathcal{V}^i\}$	The auxiliary variables copied from $\boldsymbol{x}_{j1i}$ in $\mathcal{N}^j$
$\boldsymbol{x}_{i2i}$	$\{\mathbf{f}_l^v \mid l \in \bigcup_{j\neq i} \mathcal{L}_{ij}^0, v \in \mathcal{V}^i\}$	The auxiliary variables copied from $\{\boldsymbol{x}_{i1j}\}_{j\neq i}$ in $\mathcal{N}^i$
$\boldsymbol{x}_{i3},\ \boldsymbol{x}_{i4}$	$\{\mathbf{f}_l \mid l \in \mathcal{L}^i\},\ \{\tilde{\mathbf{f}}_l \mid l \in \mathcal{L}^i\}$	The flow rate variables within $\mathcal{N}^i$, and their corresponding local copies

Table 4.1: Summary of physical meaning and the relationship for variables stored in $\mathcal{N}^i$, $i = 0 \sim K$

other. Specifically, the reformulated constraints can be expressed as

$$\mathcal{X}_0 = \{\boldsymbol{x}_0 \mid r_m \geq r_{\min},\ \forall\, m\}, \tag{4.7}$$

$$\begin{aligned}\mathcal{X}_i = \{\boldsymbol{x}_i \mid (4.6c),\ & \mathbf{1}^T\tilde{\mathbf{f}}_{l_i} \leq C_{l_i},\ \tilde{\mathbf{f}}_{l_i} \geq \mathbf{0},\ \forall\, l_i \in \mathcal{L}^i,\ \mathbf{1}^T\mathbf{f}_l \leq C_l,\ \mathbf{f}_l \geq \mathbf{0},\ \forall\, l \in \mathcal{L}_{ij}^0,\ j \neq i, \\ & \sum_{l\in \mathrm{In}(v)\cap\mathcal{L}^i} f_{l,m} + \sum_{l\in \mathrm{In}(v)\cap\mathcal{L}^0} f_{l,m}^v + 1_{v=s_m} r_m^v \\ & = \sum_{l\in \mathrm{Out}(v)\cap\mathcal{L}^i} f_{l,m} + \sum_{l\in \mathrm{Out}(v)\cap\mathcal{L}^0} f_{l,m}^v + 1_{v=d_m} r_m^v,\ \forall\, v \in \mathcal{V}^i, \forall\, m\},\ i = 1 \sim K\end{aligned} \tag{4.8}$$

In summary, problem (4.5) is equivalently reformulated as

$$\min_{\boldsymbol{x}} \quad -r_{\min} \tag{4.9a}$$

$$\text{s.t.} \quad \mathbf{A}_{ij}^i \boldsymbol{x}_i = \mathbf{A}_{ij}^j \boldsymbol{x}_j,\ \forall\, 0 \leq i < j \leq K, \tag{4.9b}$$

$$\boldsymbol{x}_i \in \mathcal{X}_i,\ i = 0 \sim K. \tag{4.9c}$$

After this reformulation, we can observe that all the constraint sets $\mathcal{X}_i$, $i = 0 \sim K$, satisfying assumption (A4) (i.e., compact and polyhedral). The coupling equality constraints between subnetworks are in the form of (4.1b). The objective function is a linear function of variable $\boldsymbol{x}_0$ while being zero function for all other variables $\boldsymbol{x}_i$, $i = 1 \sim K$, therefore satisfying assumptions (A2-A3). Furthermore, problem (4.9) is always feasible, i.e., assumption (A1) is satisfied, so it falls into the formulation (4.1).

Remark 7 *The reformulation introduced in* (4.9) *is a generalization of our previous synchronous routing algorithm in Chapter 3. The main difference is that in Chapter 3, a similar splitting is done for each node and link in the network, without modeling the physical subnetwork structure. Therefore, too many auxiliary variables are introduced. Moreover, this new reformulation enables semi-asynchronous implementation, as will be explained shortly.*

4.4 Proposed Semi-Asynchronous BSUM-M Algorithm

In this section, we formally introduce the proposed semi-asynchronous BSUM-M (Semi-BSUM-M) algorithm. We will first introduce the general algorithm description and its convergence analysis for BSUM-M with EC rule. After that, the extension that incorporates the semi-asynchronism is introduced. The specialization of the semi-asynchronous implementation for the network TE problem (4.9) will be described in Sec. 4.5.

4.4.1 The BSUM-M Algorithm with Essentially Cyclic Update Rule

We begin by simplifying the notation in problem (4.1). For each constraint of (4.1b) that couples variable $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$, i.e., $\mathbf{A}_{ij}^i \boldsymbol{x}_i + \mathbf{A}_{ij}^j \boldsymbol{x}_j = \mathbf{b}_{ij}$, $0 \leq i < j \leq K$, we write it compactly as $\tilde{\mathbf{A}}_k \boldsymbol{x} = \tilde{\mathbf{b}}_k$, where k is the constraint index; $\tilde{\mathbf{A}}_k \triangleq [\tilde{\mathbf{A}}_{0k}, \tilde{\mathbf{A}}_{1k}, \ldots, \tilde{\mathbf{A}}_{Kk}]$; $\tilde{\mathbf{A}}_{ik} = \mathbf{A}_{ij}^i$, $\tilde{\mathbf{A}}_{jk} = \mathbf{A}_{ij}^j$, and $\tilde{\mathbf{A}}_{mk} = \mathbf{0}$, $\forall\, m \neq i,\ j$; and $\tilde{\mathbf{b}}_k = \mathbf{b}_{ij}$. By denoting the total number of constraints as $J = (K+1)K/2$, problem (4.1) can be equivalently expressed

as follows

$$\min \ f(\boldsymbol{x}) = \sum_{i=0}^{K} f_i(\boldsymbol{x}_i) \tag{4.10a}$$

$$\text{s. t.} \quad \tilde{\mathbf{A}}_j \boldsymbol{x} = \tilde{\mathbf{b}}_j, \ j = 1 \sim J \tag{4.10b}$$

$$\boldsymbol{x}_i \in \mathcal{X}_i, \ i = 0 \sim K. \tag{4.10c}$$

The corresponding augmented Lagrangian function is expressed as

$$L(\boldsymbol{x}; \boldsymbol{\lambda}) = \sum_{i=0}^{K} \left[f_i(\boldsymbol{x}_i) + \sum_{j=1}^{J} \left(\langle \tilde{\mathbf{A}}_j \boldsymbol{x} - \tilde{\mathbf{b}}_j \rangle + \frac{\rho}{2} \| \tilde{\mathbf{A}}_j \boldsymbol{x} - \tilde{\mathbf{b}}_j \|^2 \right) \right]. \tag{4.11}$$

With this compact expression, we are ready to review the synchronous BSUM-M algorithm [108] for problem (4.1) in the following table.

BSUM-M with EC rule:

1: **Initialization** $\boldsymbol{x} = \mathbf{0}$, $\boldsymbol{\lambda}^0 = \mathbf{0}$, $\gamma \geq 0$, and $t = 0$

2: **Repeat**

3: **Update**

i) Primal variable $\boldsymbol{x}_i$ at node i, $i = 0 \sim K$:

$$\boldsymbol{x}_i^{t+1} = \begin{cases} \arg\min_{\boldsymbol{x}_i \in \mathcal{X}_i} L(\boldsymbol{x}_{-i}^t, \boldsymbol{x}_i; \boldsymbol{\lambda}^t) + \frac{\gamma}{2} \| \boldsymbol{x}_i - \boldsymbol{x}_i^t \|^2 & \text{if } I_{p,i}^t = 1 \\ \boldsymbol{x}_i^t & \text{if } I_{p,i}^t = 0 \end{cases},$$

ii) Dual variable $\boldsymbol{\lambda}_j$ between nodes, $j = 1 \sim J$:

$$\boldsymbol{\lambda}_j^{t+1} = \begin{cases} \boldsymbol{\lambda}_j^t + \alpha^t (\tilde{\mathbf{A}}_j \boldsymbol{x}^{t+1} - \tilde{\mathbf{b}}_j) & \text{if } I_{d,j}^t = 1 \\ \boldsymbol{\lambda}_j^t & \text{if } I_{d,j}^t = 0 \end{cases}.$$

4: $t = t + 1$

5: **Until** A desired stopping criterion is met.

In the above table, $\boldsymbol{x}_{-i}^t \triangleq [\boldsymbol{x}_0^t, \ldots, \boldsymbol{x}_{i-1}^t, \boldsymbol{x}_{i+1}^t, \ldots, \boldsymbol{x}_K^t]$; $\{I_{p,i}\}_i$ and $\{I_{d,j}\}$ are the binary variables indicating whether each of the primal variable $\boldsymbol{x}_i$ and the dual variable $\boldsymbol{\lambda}_j$ should be updated at the tth iteration; $\alpha^t > 0$ is the stepsize for updating the dual variables at iteration t. In [108], two update rules have been proposed. The first one follows the Gauss-Seidel rule which updates each $\boldsymbol{x}_i$ and $\boldsymbol{\lambda}$ sequentially, while the second

randomly selects one primal variable $\boldsymbol{x}_i$ or the dual variables $\boldsymbol{\lambda}$ to update. One can observe that BSUM-M is a synchronous algorithm: at each iteration t, in order to update $\boldsymbol{x}_i$, the computation node i requires the most current $\{\mathbf{A}_{ji}^j \boldsymbol{x}_j, \mathbf{A}_{ik}^k \boldsymbol{x}_k, \boldsymbol{\lambda}_{ji}, \boldsymbol{\lambda}_{ik}\}_{j<i,\ k>i}$ from *all* the connecting nodes. It is worth noting that if i) $\gamma = 0$; ii) a constant dual stepsize is picked, i.e., $\alpha^t = \rho$, $\forall\ t$; and iii) and Gauss-Seidel update rule is used, the synchronous BSUM-M algorithm is the same as the multi-block ADMM algorithm, see (4.3). However, as will be shown shortly, the BSUM-M with EC rule is better suited for semi-asynchronous implementation via certain reformulation. The main reason is that the multi-block nature of the considered problem can capture the asynchronism well, but to ensure convergence requires proper dual stepsize control.

In what follows, we will extend the update rule for the synchronous BSUM-M algorithm to the more general EC update rule. In particular, at the tth iteration, we require that $\sum_{i=0}^K I_{p,i}^t + \sum_{j=1}^J I_{d,j}^t = 1$, that is only *one* primal variable, $\boldsymbol{x}_i$, or *one* dual variable, $\boldsymbol{\lambda}_j$, can be updated. Furthermore, every primal and dual variable will be updated at least once for $\tau > 0$ iterations. Formally, we have the following definition.

(A5) (EC update rule) For the tth iteration $\sum_{i=0}^K I_{p,i}^t + \sum_{j=1}^J I_{d,j}^t = 1$, $\forall\ t$. Moreover, there exists a period $\tau \geq 1$ during which each primal and dual variable is updated at least once, i.e. for every $T \geq 0$,

$$\sum_{t=T+1}^{t+\tau} I_{p,i}^t \geq 1, \quad \sum_{t=T+1}^{t+\tau} I_{d,j}^t \geq 1,\ \forall\ i,\ j. \tag{4.12}$$

The EC update rule is the generalization of the more restricted Gauss-Seidel one, which has essentiality $\tau = 1 + K + J$ and a fixed update sequence. Also, it can be any deterministic update sequence following the definition (A5) without any underlying probability model as assumed in the randomized update rule. We first analyze the convergence property of the BSUM-M algorithm with EC update rule, and it is shown in the following theorem.

Theorem 5 *Suppose the assumptions (A1)–(A5) hold. Assume the α^t satisfies the following stepsize rule:*

$$\sum_{t=1}^{\infty} \alpha^t = \infty, \quad \lim_{t\to\infty} \alpha^t = 0. \tag{4.13}$$

If

$$\gamma + \frac{\rho\lambda_{\min}(\sum_{j=1}^{J}\tilde{\mathbf{A}}_{ij}^{T}\tilde{\mathbf{A}}_{ij})}{2} > 0,\ i = 0 \sim K, \tag{4.14}$$

then $\lim_{t\to\infty}\|\boldsymbol{x}^{t+1} - \boldsymbol{x}^{t}\| = 0$ *and* $\lim_{t\to\infty}\|\tilde{\mathbf{A}}_j\boldsymbol{x}^{t} - \tilde{\mathbf{b}}_j\| = 0$, $j = 1 \sim J$. *Furthermore, every limit point of* $\{\boldsymbol{x}^{t}, \boldsymbol{\lambda}^{t}\}$ *generated by the BSUM-M algorithm with essential cyclic rule is a primal and dual optimal solution.*

The proof of Theorem 5 is relegated to the appendix D. Note that by examining the analysis details, the convergence analysis can be extended to incorporate some non-smooth objective functions. Specifically, each objective function $f_i(\boldsymbol{x}_i)$, $\forall\ i$, can be replaced by

$$f_i(\boldsymbol{x}_i) + w_i\|\boldsymbol{x}_i\|_1 + \sum_{m} w_{i,m}\|\boldsymbol{x}_{i,m}\|_2,$$

where $w_i \geq 0$ and $w_{i,m} \geq 0$ are some constants and $\boldsymbol{x}_i = (\ldots, \boldsymbol{x}_{i,m}, \ldots)$ is a partition of $\boldsymbol{x}_i$ with m being the partition index. This is due to the fact that the critical local error bound property used in the proof of Theorem 5 still holds; cf. (D.14) in Appendix D. Moreover, the stepsize rule of α^t given in (4.13) indicates that it should be sufficiently small in the end. This explains why the direct application of multi-block ADMM with fixed stepsize ρ may not work. Furthermore, the extra γ term should satisfy the condition (4.14), which ensures the augmented Lagrangian function plus the extra quadratic term with coefficient γ is strongly convex for each primal variable. Therefore, after each primal variable update, the value of the augmented Lagrangian function can be strictly decreased.

In the next subsection, we will properly reformulate the considered problem (4.1) to incorporate the different computation and communication delays between nodes. After the reformulation, we will show that BSUM-M with EC rule results in the desired semi-asynchronous property.

4.4.2 The Proposed Semi-Asynchronous BSUM-M Algorithm

In this section, we formally introduce the proposed algorithm suitable for the semi-asynchronous network model. In contrast to the partially asynchronous model [38], our

semi-asynchronous model does not allow out-of-sequence information exchange nor communication failure. However, the system can still tolerate communication/computation delays: each node is able to process the updates from other nodes as soon as they arrive instead of waiting the updates from all other nodes while information from the rest of the nodes can be outdated. To achieve this, the key is to add a few "buffer" variables at each node, which decouple the linear equality constraints (4.1b) across the nodes. Effectively, from a particular node's perspective, these variables record the latest states of the rest of the nodes.

Let us denote the new set of auxiliary variables $\bar{\boldsymbol{x}}_{ij}^i$ and $\bar{\boldsymbol{x}}_{ij}^j$, $0 \leq i < j \leq K$, which are stored in the ith and jth node, respectively. Each $\bar{\boldsymbol{x}}_{ij}^i$ and $\bar{\boldsymbol{x}}_{ij}^j$ are, respectively, defined to be equal to $\mathbf{A}_{ij}^i \boldsymbol{x}_i$ and $\mathbf{A}_{ij}^j \boldsymbol{x}_j$. Thus the linear equality constraints (4.1b) can be rewritten as

$$\mathbf{A}_{ij}^i \boldsymbol{x}_i = \bar{\boldsymbol{x}}_{ij}^i,\ \mathbf{A}_{ij}^j \boldsymbol{x}_j = \bar{\boldsymbol{x}}_{ij}^j,\ \bar{\boldsymbol{x}}_{ij}^i + \bar{\boldsymbol{x}}_{ij}^j = \mathbf{b}_{ij},\ 0 \leq i < j \leq K. \tag{4.15}$$

Using these new auxiliary variables, problem (4.1) can be equivalently reformulated as

$$\min_{\boldsymbol{x}} \quad f(\boldsymbol{x}) = \sum_{i=0}^{K} f_i(\boldsymbol{x}_i) \tag{4.16a}$$

$$\text{s.t.} \quad (4.15) \text{ and } \boldsymbol{x}_i \in \mathcal{X}_i,\ i = 0 \sim K. \tag{4.16b}$$

The corresponding augmented Lagrangian function is expressed as,

$$\begin{aligned}\bar{L}(\boldsymbol{x}, \bar{\boldsymbol{x}}; \bar{\boldsymbol{\lambda}}) = & \sum_{0 \leq i < j \leq K} \underbrace{\left[\langle \bar{\boldsymbol{x}}_{ij}^i + \bar{\boldsymbol{x}}_{ij}^j - \mathbf{b}_{ij}, \bar{\boldsymbol{\lambda}}_{ij} \rangle + \frac{\rho}{2} \| \bar{\boldsymbol{x}}_{ij}^i + \bar{\boldsymbol{x}}_{ij}^j - \mathbf{b}_{ij} \|^2 \right]}_{\triangleq \bar{L}_{ij}(\bar{\boldsymbol{x}}_{ij}^i, \bar{\boldsymbol{x}}_{ij}^j; \bar{\boldsymbol{\lambda}}_{ij})} \\ & + \sum_{i=0}^{K} \underbrace{\left[f_i(\boldsymbol{x}_i) + \sum_{j<i} \langle \mathbf{A}_{ji}^i \boldsymbol{x}_i - \bar{\boldsymbol{x}}_{ji}^i, \bar{\boldsymbol{\lambda}}_{ji}^i \rangle + \frac{\rho}{2} \| \mathbf{A}_{ji}^i \boldsymbol{x}_i - \bar{\boldsymbol{x}}_{ji}^i \|^2 \right.}_{} \\ & \qquad \underbrace{\left. + \sum_{k>i} \langle \mathbf{A}_{ik}^i \boldsymbol{x}_i - \bar{\boldsymbol{x}}_{ik}^i, \bar{\boldsymbol{\lambda}}_{ik}^i \rangle + \frac{\rho}{2} \| \mathbf{A}_{ik}^i \boldsymbol{x}_i - \bar{\boldsymbol{x}}_{ik}^i \|^2 \right]}_{\triangleq \bar{L}_i(\boldsymbol{x}_i, \{\bar{\boldsymbol{x}}_{ji}^i\}_{j<i}, \{\bar{\boldsymbol{x}}_{ik}^i\}_{k>i}; \{\bar{\boldsymbol{\lambda}}_{ji}^i\}_{j<i}, \{\bar{\boldsymbol{\lambda}}_{ik}^i\}_{k>i})}\end{aligned} \tag{4.17}$$

where $\bar{\boldsymbol{x}} \triangleq \{\bar{\boldsymbol{x}}_{ij}^i, \bar{\boldsymbol{x}}_{ij}^j \mid 0 \leq i < j \leq K\}$, and $\bar{\boldsymbol{\lambda}} \triangleq \{\bar{\boldsymbol{\lambda}}_{ij}^i, \bar{\boldsymbol{\lambda}}_{ij}^j, \bar{\boldsymbol{\lambda}}_{ij} \mid 0 \leq i < j \leq K\}$ is the set of dual variable for linear equality constraints of (4.15). Notice that only the extra

auxiliary variables $\bar{\boldsymbol{x}}_{ij}^i$ and $\bar{\boldsymbol{x}}_{ij}^j$, which are stored in different nodes, are coupled to each other through $\bar{L}_{ij}(\bar{\boldsymbol{x}}_{ij}^i, \bar{\boldsymbol{x}}_{ij}^j; \bar{\boldsymbol{\lambda}}_{ij})$, $0 \leq i < j \leq K$. On the other hand, the augmented Lagrangian $\bar{L}(\boldsymbol{x}, \bar{\boldsymbol{x}}; \bar{\boldsymbol{\lambda}})$ is separable over $\{\boldsymbol{x}_i\}_{i=0}^K$. By exploiting this separability property, we propose the Semi-BSUM-M algorithm, which is summarized in Table 4.2.

Let us explain the algorithm in detail for the tth iteration of node i. First, let us use the binary variables $I_{ji}^{i,t}$ and $I_{ik}^{i,t}$ to respectively denote whether auxiliary variables and the dual variables $\{\bar{\boldsymbol{x}}_{ji}^i, \bar{\boldsymbol{\lambda}}_{ji}\}$ and $\{\bar{\boldsymbol{x}}_{ik}^i, \boldsymbol{\lambda}_{ik}\}$, $j < i$ and $k > i$, should be updated. Node i receives the updated information only from a subset of other nodes denoted as $\mathcal{S}_{t,i} \subseteq \{0, \ldots, K\}$. In particular, new information $\bar{\boldsymbol{x}}_{ji}^j$ and $\bar{\boldsymbol{x}}_{ik}^k$ is received from the "active" nodes $j, k \in \mathcal{S}_{t,i}$, $j < i$ and $k > i$. This information is used to update the local variables and the dual variables $\{\bar{\boldsymbol{x}}_{ji}^i, \bar{\boldsymbol{\lambda}}_{ji}\}$ and $\{\bar{\boldsymbol{x}}_{ik}^i, \bar{\boldsymbol{\lambda}}_{ik}\}$, cf. (4.18) and (4.19), so $I_{ji}^{i,t} = 1$ and $I_{ik}^{i,t} = 1$. On the other hand, the variables $\{\bar{\boldsymbol{x}}_{j'i}^i, \bar{\boldsymbol{\lambda}}_{j'i}\}$ and $\{\bar{\boldsymbol{x}}_{ik'}^i, \bar{\boldsymbol{\lambda}}_{ik'}\}$ related to the "inactive" nodes $j', k' \in \mathcal{S}_{t,i}^c$, $j' < i$ and $k' > i$, remain the same until these nodes become "active", and $I_{j'i}^{i,t}$ and $I_{ik'}^{i,t}$ are zero. After the updates of auxiliary variables, the local variable $\boldsymbol{x}_i$ will be updated, so $I_{p,i}^t = 1$. Furthermore, semi-asynchronous model requires the mild bounded delay assumption at each node, i.e., the outdatedness between nodes cannot exceed a fixed upper bound. To guarantee this assumption, each node needs to keep track of the outdatedness of the local variables, and will stop updating if there is no update from a specific node for a predetermined upper bound τ^{semi}. With the bounded delay assumption, for each node all its direct neighbors should be updated at least once within τ^{semi} local iterations, and the update sequence is strongly related to the subset $\mathcal{S}_{t,i}$. Hence, we observe that the proposed Semi-BSUM-M algorithm indeed belongs to the BSUM-M with EC rule, because there exists a bounded time interval in which all variable will be updated at least once (whose bound is related to τ^{semi}).

Second, the additional tolerance of the semi-asynchronism for Semi-BSUM-M can be obtained as follows. Combining with the fact that the update of local variable $\boldsymbol{x}_i$ (cf. (4.20)) only depends on the auxiliary variables $\{\bar{\boldsymbol{x}}_{ji}^i\}_{j<i}$ and $\{\bar{\boldsymbol{x}}_{ik}^k\}_{k>i}$, it follows that these extra auxiliary variables serve as a "buffer" which stores the latest information from each of the connected node. Moreover, the node i only sends out the updated information to the active nodes at its tth iteration, triggering the next round of buffer variable updates at those active nodes. After that, the new snapshot of the buffer variable will be sent back to node i.

Semi-BSUM-M: Steps for node i, $i = 0 \sim K$

1: **Initialization** $\boldsymbol{x}^0 = \mathbf{0}$, $\bar{\boldsymbol{x}}^0 = \mathbf{0}$, $\bar{\boldsymbol{\lambda}}^0 = \mathbf{0}$, $\gamma > 0$, $\{\alpha^t\}$, $\tau^{\text{semi}} > 0$, and $t = 0$

2: **Repeat**

3: **Wait** until receiving $\{\hat{\boldsymbol{x}}^j_{ji}, \hat{\boldsymbol{\lambda}}_{ji}\}_{j<i}$ and $\{\hat{\boldsymbol{x}}^k_{ik}, \hat{\boldsymbol{\lambda}}_{ik}\}_{k>i}$ from node j, $k \in \mathcal{S}_{t,i}$. For $j', k' \in \mathcal{S}^c_{t,i}$, $j' < i$, $k' > i$, $\{\tau^t_{j'i}, \tau^t_{ik'}\} < \tau^{\text{semi}}$.

4: **Update** For $\forall\, j < i$:

If $j \in \mathcal{S}^c_{t,i}$, i.e., $I^{i,t}_{ji} = 0$: $\bar{\boldsymbol{x}}^{j,t+1}_{ji} = \bar{\boldsymbol{x}}^{j,t}_{ji}$, $\bar{\boldsymbol{\lambda}}^{t+1}_{ji} = \bar{\boldsymbol{\lambda}}^t_{ji}$, and $\tau^{t+1}_{ji} = \tau^t_{ji} + 1$

If $j \in \mathcal{S}_{t,i}$, i.e., $I^{i,t}_{ji} = 1$: $\bar{\boldsymbol{x}}^{j,t+1}_{ji} = \hat{\boldsymbol{x}}^j_{ji}$, $\bar{\boldsymbol{\lambda}}^{t+1}_{ji} = \hat{\boldsymbol{\lambda}}_{ji}$, $\tau^{t+1}_{ji} = 0$

$$\begin{aligned}\bar{\boldsymbol{x}}^{i,t+1}_{ji} = &\arg\min_{\bar{\boldsymbol{x}}^i_{ji}} \langle \mathbf{A}^i_{ji}\boldsymbol{x}^t_i - \bar{\boldsymbol{x}}^i_{ji}, \bar{\boldsymbol{\lambda}}^{i,t+1}_{ji}\rangle + \frac{\rho}{2}\|\mathbf{A}^i_{ji}\boldsymbol{x}^t_i - \bar{\boldsymbol{x}}^i_{ji}\|^2 \\ &+ \bar{L}_{ji}(\bar{\boldsymbol{x}}^{j,t+1}_{ji}, \bar{\boldsymbol{x}}^i_{ji}; \bar{\boldsymbol{\lambda}}^{t+1}_{ij}) + \frac{\gamma}{2}\|\bar{\boldsymbol{x}}^i_{ji} - \bar{\boldsymbol{x}}^{i,t}_{ji}\|^2 \qquad (4.18)\end{aligned}$$

$$\bar{\boldsymbol{\lambda}}^{t+1}_{ji} = \bar{\boldsymbol{\lambda}}^{t+1}_{ji} + \alpha^t(\bar{\boldsymbol{x}}^{j,t+1}_{ji} + \bar{\boldsymbol{x}}^{i,t+1}_{ji} - \mathbf{b}_{ji})$$

Send $\{\bar{\boldsymbol{x}}^{i,t+1}_{ji}, \bar{\boldsymbol{\lambda}}^{t+1}_{ji}\}_{k>i}$ to node $j \in \mathcal{S}_{t,i}$

5: **Update** For $\forall\, k > i$:

If $k \in \mathcal{S}^c_{t,i}$, i.e., $I^{i,t}_{ik} = 0$: $\bar{\boldsymbol{x}}^{k,t+1}_{ik} = \bar{\boldsymbol{x}}^{k,t}_{ik}$, $\bar{\boldsymbol{\lambda}}^{t+1}_{ik} = \bar{\boldsymbol{\lambda}}^t_{ik}$, and $\tau^{t+1}_{ik} = \tau^t_{ik} + 1$

If $k \in \mathcal{S}_{t,i}$, i.e., $I^{i,t}_{ik} = 1$: $\bar{\boldsymbol{x}}^{k,t+1}_{ik} = \hat{\boldsymbol{x}}^k_{ik}$, $\bar{\boldsymbol{\lambda}}^{t+1}_{ik} = \hat{\boldsymbol{\lambda}}_{ik}$, $\tau^{t+1}_{ik} = 0$

$$\begin{aligned}\bar{\boldsymbol{x}}^{i,t+1}_{ik} = &\arg\min_{\bar{\boldsymbol{x}}^i_{ik}} \langle \mathbf{A}^i_{ik}\boldsymbol{x}^t_i - \bar{\boldsymbol{x}}^i_{ik}, \bar{\boldsymbol{\lambda}}^{i,t+1}_{ik}\rangle + \frac{\rho}{2}\|\mathbf{A}^i_{ik}\boldsymbol{x}^t_i - \bar{\boldsymbol{x}}^i_{ik}\|^2 \\ &+ \bar{L}_{ik}(\bar{\boldsymbol{x}}^i_{ik}, \bar{\boldsymbol{x}}^{k,t+1}_{ik}; \bar{\boldsymbol{\lambda}}^{t+1}_{ik}) + \frac{\gamma}{2}\|\bar{\boldsymbol{x}}^i_{ik} - \bar{\boldsymbol{x}}^{i,t}_{ik}\|^2 \qquad (4.19)\end{aligned}$$

$$\bar{\boldsymbol{\lambda}}^{t+1}_{ik} = \bar{\boldsymbol{\lambda}}^{t+1}_{ik} + \alpha^t(\bar{\boldsymbol{x}}^{i,t+1}_{ik} + \bar{\boldsymbol{x}}^{k,t+1}_{ik} - \mathbf{b}_{ik})$$

Send $\{\bar{\boldsymbol{x}}^{i,t+1}_{ik}, \bar{\boldsymbol{\lambda}}^{t+1}_{ik}\}_{k>i}$ to node $k \in \mathcal{S}_{t,i}$

6: **Update** $(I^t_{p,i} = I^t_{d,i} = 1)$ For $\forall\, j < i < k$:

$$\boldsymbol{x}^{t+1}_i = \arg\min_{\boldsymbol{x}_i \in \mathcal{X}_i} \bar{L}_i(\boldsymbol{x}_i, \{\bar{\boldsymbol{x}}^{i,t+1}_{ji}\}_{j<i}, \{\bar{\boldsymbol{x}}^{i,t+1}_{ik}\}_{k>i}; \{\bar{\boldsymbol{\lambda}}^{i,t}_{ji}\}_{j<i}, \{\bar{\boldsymbol{\lambda}}^{i,t}_{ik}\}_{k>i}) + \frac{\gamma}{2}\|\boldsymbol{x}_i - \boldsymbol{x}^t_i\|^2, \qquad (4.20)$$

$$\bar{\boldsymbol{\lambda}}^{i,t+1}_{ji} = \begin{cases} \bar{\boldsymbol{\lambda}}^{i,t}_{ji} + \alpha^t(\mathbf{A}^i_{ji}\boldsymbol{x}^{t+1}_i - \bar{\boldsymbol{x}}^{i,t+1}_{ji}) & ,\ j \in \mathcal{S}_{t,i} \\ \bar{\boldsymbol{\lambda}}^{i,t}_{ji} & ,\ j \in \mathcal{S}^c_{t,i} \end{cases},$$

$$\bar{\boldsymbol{\lambda}}^{i,t+1}_{ik} = \begin{cases} \bar{\boldsymbol{\lambda}}^{i,t}_{ik} + \alpha^t(\mathbf{A}^i_{ik}\boldsymbol{x}^{t+1}_i - \bar{\boldsymbol{x}}^{i,t+1}_{ik}) & ,\ k \in \mathcal{S}_{t,i} \\ \bar{\boldsymbol{\lambda}}^{i,t}_{ik} & ,\ k \in \mathcal{S}^c_{t,i} \end{cases}$$

7: $t = t + 1$

8: **Until** A given stopping criterion is met

Table 4.2: The update procedure of the proposed Semi-BSUM-M algorithm.

To give a more concrete illustration about why the asynchronism between nodes can be handled via the help of the extra buffer variables, in Fig.4.5, we provide an example consisting of three nodes indexed by i, j, and k, $j < i < k$. The two subfigures in the figure represent two consecutive iterations. In Fig.4.5 (a), some of the buffer variables are updated within node i and k while others are kept fixed. As shown in the pseudo code of Semi-BSUM-M algorithm, the original variables $\boldsymbol{x}_i$ and $\boldsymbol{x}_k$, respectively, in node i and k will be updated immediately after the updates of the buffer variables, i.e., $\bar{\boldsymbol{x}}_{ji}^i$ and $\{\bar{\boldsymbol{x}}_{jk}^k, \bar{\boldsymbol{x}}_{ik}^k\}$, respectively. In node j no update occurs since all buffer variable are updated at the other side of the connecting link. When we proceed to Fig.4.5 (b), the buffer variable update in node i, i.e., $\bar{\boldsymbol{x}}_{ji}^i$ has been completed, and the new information has been transmitted to node j. The buffer variables $\bar{\boldsymbol{x}}_{ji}^j$ at the other side start updating, and the local variables $\boldsymbol{x}_j$ will be recomputed after the buffer variables have been renewed. In node k, during the two consecutive iterations in Fig.4.5, the buffer variables $\bar{\boldsymbol{x}}_{jk}^k$ and $\bar{\boldsymbol{x}}_{ik}^k$ and the local variable $\boldsymbol{x}_k$ have not finished updating yet while the local variables of the other two nodes have changed. This illustrates why with the help of the introduced buffer variables, the proposed approach allows each node to process its variables at different speeds. The consistency of the variables is ensured in the limit as the algorithm converges.

From the previous discussion we conclude that despite the exisitence of asynchronism between nodes, all variables are updated according to the most up-to-date information due to the existence of buffer variables. Since the proposed algorithm is BSUM-M with EC rule, by Theorem 5, we have the following convergence property for Semi-BSUM-M algorithm

Corollary 1 *For Semi-BSUM-M algorithm, if i) the delay upper bound τ^{semi} is finite; ii) assumptions (A1)-(A4) hold; and iii) the parameter α^t and γ satisfy the conditions as those in Theorem 5, then every limit point of $\{\boldsymbol{x}^t, \boldsymbol{\lambda}^t\}$ generated by Semi-BSUM-M algorithm is a primal and dual optimal solution of problem* (4.1).

Before closing this subsection, wecomment on the choice of the parameter ρ. To reduce the communication overhead between nodes with fewer number of iterations to achieve convergence, one can apply different ρ_{ij}^k, $0 \leq i < j \leq K$ and $k = 1 \sim 3$, for each individual constraint of (4.15). Specifically, the primal residuals $r_{ij}^{k,t}$ and dual residuals

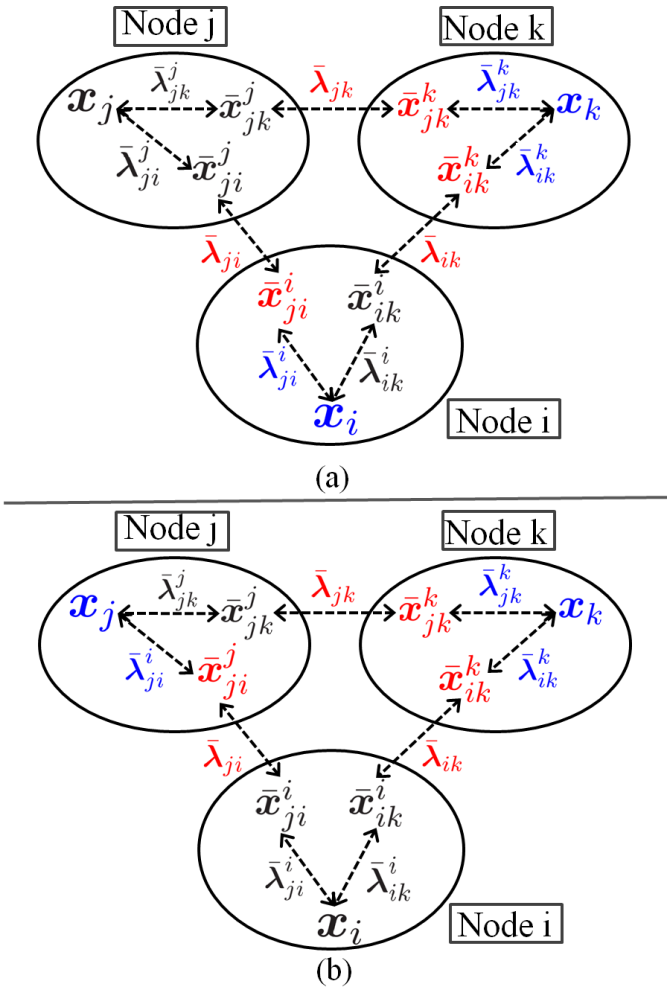


Figure 4.5: An illustrating example consists of three nodes, $j < i < k$, for the proposed Semi-BSUM-M algorithm over two consecutive iterations. (a) corresponds to the first iteration, and (b) is the second. The blue color indicates the original variables is updating. The red color represents the extra buffer variables is updating. The black means the variables remain fixed during the current time.

$s_{ij}^{k,t}$ at the tth iteration can be derived similarly to those in [39, Chap.3.3] with minor modification to incorporate the time varying effect of $\{\alpha^t\}$ as follows:

$$\begin{aligned}
& r_{ij}^{1,t} \triangleq \mathbf{A}_{ij}^i \boldsymbol{x}_i^t - \bar{\boldsymbol{x}}_{ij}^{i,t}, r_{ij}^{2,t} \triangleq \mathbf{A}_{ij}^j \boldsymbol{x}_j^t - \bar{\boldsymbol{x}}_{ij}^{j,t},\ r_{ij}^{3,t} \triangleq \bar{\boldsymbol{x}}_{ij}^{i,t} + \bar{\boldsymbol{x}}_{ij}^{j,t} - \mathbf{b}_{ij}, \\
& s_{ij}^{1,t} \triangleq \mathbf{A}_{ij}^i (\rho_{ij}^1 \boldsymbol{x}_i^{t-1} - \boldsymbol{x}_i^t) + (\alpha^{t-1} - \rho_{ij}^1) \bar{\boldsymbol{x}}_{ij}^{i,t}, \\
& s_{ij}^{2,t} \triangleq \mathbf{A}_{ij}^j (\rho_{ij}^2 \boldsymbol{x}_j^{t-1} - \boldsymbol{x}_j^t) + (\alpha^{t-1} - \rho_{ij}^2) \bar{\boldsymbol{x}}_{ij}^{j,t}, \\
\text{and} \quad & s_{ij}^{3,t} \triangleq (\alpha^{t-1} - \rho_{ij}^3)(\bar{\boldsymbol{x}}_{ij}^{j,t} - \mathbf{b}_{ij}) + \alpha^{t-1} \bar{\boldsymbol{x}}_{ij}^{i,t} - \rho_{ij}^3 \bar{\boldsymbol{x}}_{ij}^{i,t-1}.
\end{aligned}$$

With these primal/dual residuals, each ρ_{ij}^k is adaptively adjusted by the following mechanism:

$$\rho_{ij}^{k,t+1} = \begin{cases} \tau^{\text{incr}} \rho_{ij}^{k,t} & \text{if } \|r_{ij}^{k,t}\| > \mu \|s_{ij}^{k,t}\| \\ \rho_{ij}^{k,t} / \tau^{\text{decr}} & \text{if } \|s_{ij}^{k,t}\| > \mu \|r_{ij}^{k,t}\| \\ \rho_{ij}^{k,t} & \text{otherwise} \end{cases}, \tag{4.21}$$

where μ, τ^{incr}, and τ^{decr} are predetermined parameters; ρ_{ij}^1 and ρ_{ij}^2 are updated in node i and node j, respectively; and ρ_{ij}^3 is updated in node i with $i < j$.

4.5 Application to the Network TE Problem

While the generic Semi-BSUM-M algorithm proposed in the previous section can effectively handle the asynchronism, it does at the same time require solving a convex problem (4.20) exactly, which can be expensive. Indeed, when specializing to the TE problem (4.9), problem (4.20) involves the difficult constraint $\mathcal{X}_i$ (4.8) which consists of both the flow conservation constraints and the capacity constraints. In this section, we specialize the proposed Semi-BSUM-M algorithm to the network TE problem (4.9) in a way that further reduces computational complexity. Moreover, we will account for the existence of the master node, and use this fact to further simplify the implementation.

Specifically, we shift the flow rate control for the links across different subnetworks, i.e., $\mathcal{L}^0$, to the master node, which we denote as NC 0. The constraint set at NC 0 is therefore changed to

$$\mathcal{X}_0 = \{\boldsymbol{x}_0 \mid r_m \geq r_{\min},\ \forall\, m,\ \mathbf{1}^T \mathbf{f}_l \leq C_l,\ \mathbf{f}_l \geq 0,\ \forall\, l \in \mathcal{L}^0\}. \tag{4.22}$$

Moreover, the corresponding local update procedure (4.20) at NC 0 is decomposable across links, and can be performed in closed-form as in (3.22). This modification allows each subnetwork to communicate with NC 0 instead of communicating among themselves.

Similarly as before, to handle asynchronism between NC i, $i = 1 \sim K$ and NC 0, we need to introduce "buffer" variables for NC 0, denoted as $\tilde{f}_l^{s_l}$ and $\tilde{f}_l^{d_l}$, $\forall\, l \in \mathcal{L}^0$. Moreover, since the distributed NCs no longer communicate with each other, only one set of buffer variable per distributed NC is enough. In other words, to reduce the number of the auxiliary variables, the buffer variables are not introduced at each distributed NC. Instead, the original variables with flow conservation constraints will also serve as the role of the buffer variables. By doing this, we reduce the number of auxiliary variables and the computational effort to update them (cf. (4.18) and (4.19)). Similar reformulation can be applied to the variables for the commodity rates as well. The resulting relationship across variables is shown in Fig.4.6.

Furthermore, within each subnetwork i, $i = 1 \sim K$, the variables $\boldsymbol{x}_i$ can be split into two independent variable sets, i.e., $\boldsymbol{x}_i = [(\boldsymbol{x}_i^1)^T, (\boldsymbol{x}_i^2)^T]^T$ where $\boldsymbol{x}_i^1 \triangleq [(\boldsymbol{x}_{i0})^T, (\{\boldsymbol{x}_{i2j}\}_j)^T, (\boldsymbol{x}_{i3})^T]^T$ and $\boldsymbol{x}_i^2 \triangleq \boldsymbol{x}_{i4}$. By using this expression for $\boldsymbol{x}_i$, the relationship $\boldsymbol{x}_{i3} = \boldsymbol{x}_{i4}$ indicates that there exist matrices $\mathbf{A}_i^1 = [\mathbf{0}\ \mathbf{0}\ \mathbf{I}]$ and $\mathbf{A}_i^2 = \mathbf{I}$ satisfying $\mathbf{A}_i^1 \boldsymbol{x}_i^1 = \mathbf{A}_i^2 \boldsymbol{x}_i^2$. Each of $\boldsymbol{x}_i^1$ and $\boldsymbol{x}_i^2$ has its own constraint set split from $\mathcal{X}_i$, which is expressed as follows,

$$\begin{aligned}
\mathcal{X}_i^1 = \{\boldsymbol{x}_i^1 \mid & \sum_{l \in \mathrm{In}(v) \cap \mathcal{L}^i} f_{l,m} + \sum_{l \in \mathrm{In}(v) \cap \mathcal{L}^0} f_{l,m}^v + 1_{v=s_m} r_m^v \\
& = \sum_{l \in \mathrm{Out}(v) \cap \mathcal{L}^i} f_{l,m} + \sum_{l \in \mathrm{Out}(v) \cap \mathcal{L}^0} f_{l,m}^v + 1_{v=d_m} r_m^v,\ \forall\, v \in \mathcal{V}^i, \forall\, m\}, \\
\mathcal{X}_i^2 = \{\boldsymbol{x}_i^2 \mid & \mathbf{1}^T \tilde{\mathbf{f}}_{l_i} \le C_{l_i},\ \tilde{\mathbf{f}}_{l_i} \ge \mathbf{0},\ \forall\, l_i \in \mathcal{L}^i\}.
\end{aligned}$$

In the following, we will exploit this expression for $\boldsymbol{x}_i$ and the fact that the proposed Semi-BSUM-M is the BSUM-M with the EC update rule and with the help of the extra auxiliary buffer variables. Specifically, the two independent variable sets of $\boldsymbol{x}_i$ are coupled to each other through a linear equality constraint as in BSUM-M (4.10b). Thus, instead of jointly updating them as in (4.20), we can sequentially update them so that the number of variable sets within node i is increased to two. The update procedure (4.20) in each node i is therefore replaced by the following one without affecting the

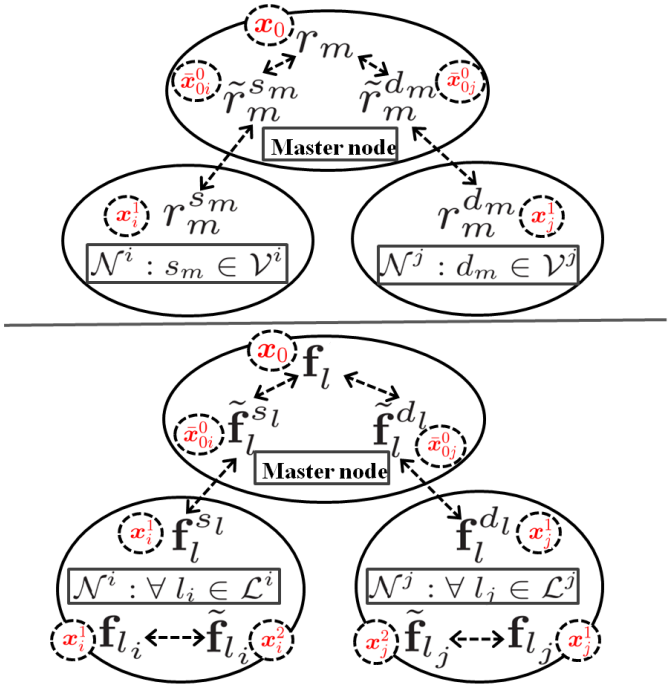


Figure 4.6: The relationship across the (auxiliary) variables of semi-asynchronous implementation for network TE problem. The variables connected by dash lines should be equal to each other, and each belongs to the variable group inside the closest dash circle.

convergence property

$$\boldsymbol{x}_i^{1,t+1} = \arg \min_{\boldsymbol{x}_i^1 \in \mathcal{X}_i^1} \langle \mathbf{A}_i^1 \boldsymbol{x}_i^1 - \mathbf{A}_i^2 \boldsymbol{x}_i^{2,t}, \bar{\boldsymbol{\lambda}}_i^t \rangle + \frac{\rho}{2} \|\mathbf{A}_i^1 \boldsymbol{x}_i^1 - \mathbf{A}_i^2 \boldsymbol{x}_i^{2,t}\|^2 + L_i([\boldsymbol{x}_i^1, \boldsymbol{x}_i^{2,t}], \{\bar{\boldsymbol{x}}_{ji}^{i,t+1}\}_{j<i}, \{\bar{\boldsymbol{x}}_{ik}^{i,t+1}\}_{k>i}; \{\boldsymbol{\lambda}_{ji}^{i,t}\}_{j<i}, \{\boldsymbol{\lambda}_{ik}^{i,t}\}_{k>i}) \quad (4.23a)$$

$$\boldsymbol{x}_i^{2,t+1} = \arg \min_{\boldsymbol{x}_i^2 \in \mathcal{X}_i^2} \langle \mathbf{A}_i^1 \boldsymbol{x}_i^{1,t+1} - \mathbf{A}_i^2 \boldsymbol{x}_i^2, \bar{\boldsymbol{\lambda}}_i^t \rangle + \frac{\rho}{2} \|\mathbf{A}_i^1 \boldsymbol{x}_i^{1,t+1} - \mathbf{A}_i^2 \boldsymbol{x}_i^2\|^2 \quad (4.23b)$$

$$\bar{\boldsymbol{\lambda}}_i^{t+1} = \bar{\boldsymbol{\lambda}}_i^t + \alpha^t (\mathbf{A}_i^1 \boldsymbol{x}_i^{1,t+1} - \mathbf{A}_i^2 \boldsymbol{x}_i^{2,t+1}). \quad (4.23c)$$

where the dual variable for constraint $\mathbf{A}_i^1 \boldsymbol{x}_i^1 = \mathbf{A}_i^2 \boldsymbol{x}_i^2$ is denoted as $\bar{\boldsymbol{\lambda}}_i$. Note that in this new update rule (4.23), $\gamma = 0$ since the condition on γ (4.14) is satisfied. The problem (4.23a) is in the traditional network optimization problem with quadratic objective function and flow conservation constraints, and it can be efficiently solved by, e.g., the RELAX code [70]. The update procedure (4.23b) is decomposable over each link, and each of them can be solved in closed-form as in (3.22). Hence, each update step in Semi-BSUM-M algorithm for the network TE problem has efficient solution, and the computational complexity is effectively reduced.

4.6 Numerical Experiments

In this section, we report some numerical results on the performance of the proposed Semi-BSUM-M algorithm for solving the network TE problem discussed in Sec.4.3. We consider a hierarchical network with 126 network nodes. These network nodes are partitioned into 9 subnetworks with 306 directed links within these subnetworks and 100 directed links connecting the subnetworks. Each subnetwork is controlled by a local NC, i.e., $K = 9$, and there is one central NC 0. The topology and the connectivity of this hierarchical network are shown in Fig.4.7. The link capacities are generated uniformly randomly in each simulation sample. In particular, links within each subnetwork have capacity distributed according to Uniform$[50, 100]$ (MBits/s) and links between each subnetwork are distributed according to Uniform$[20, 50]$ (MBits/s). The source and the destination nodes of each data flow are randomly selected from network nodes, and all simulation results are averaged over 200 randomly selected data flow pairs and link capacities. The stepsize for updating the dual variables is set as $\alpha^t = 100/(\sqrt{t}+100)$. Let $\gamma = 0$ and each ρ is initialized as 0.0005 and is adaptively updated according to (4.21)

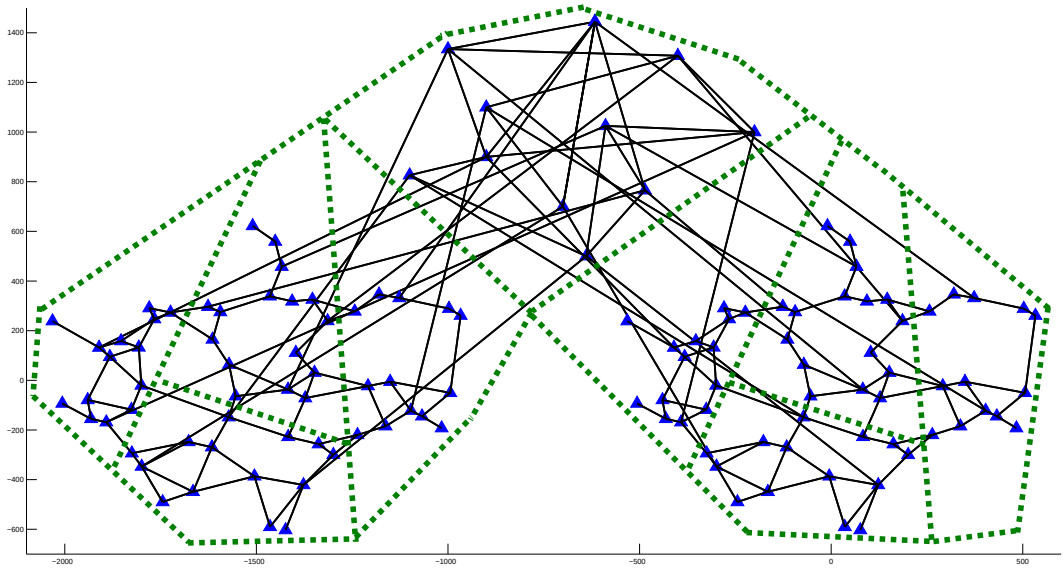


Figure 4.7: The considered network topology with 9 subnetworks.

with $\mu = 100$, $\tau^{\text{incr}} = \tau^{\text{decr}} = 1.2$ with maximum and minimum values of ρ being 1 and 0.0005. The bounds for the maximum delay τ^{semi} is set to 10. The cardinality of the active set at NC 0 should satisfy $|\mathcal{S}_{t,0}| \geq 1$, and at NC i, $i = 1 \sim K$, only information from NC 0 is received with $|\mathcal{S}_{t,i}| = 1$, $\forall\, t$. Both the synchronous and the semi-asynchronous distributed algorithms are implemented using the Open MPI package with 10 computational nodes on a SunFire X4600 server with AMD Opteron 8356 2.3GHz CPUs, and each computational node serves as one NC. No artificial communication delay between nodes is imposed.

In the experiment, we compare the proposed semi-asynchronous algorithm for formulation (4.16) with the synchronous BSUM-M algorithm where NC 0 updates its local variables only when the latest information from all subnetworks are available. Three performance metrics are used. The relative error in objective and the maximum constraint violation are, respectively, defined as $|r_{\min}^t - r_{\min}^{\text{optimal}}|/r_{\min}^{\text{optimal}}$ and the maximum $|x - x^{\text{local}}|/\max\{1, x\}$ over all variables where x (resp. x^{local}) is the original variable (resp. local auxiliary one). The successive improvement is defined as $|r_{\min}^t - r_{\min}^{t-1}|/r_{\min}^{t-1}$. Since the considered hierarchical network traffic engineering problem (4.5) is a linear

program (LP), we also compare its efficiency with the commercial LP solver, Gurobi [121]. In Table 4.3, the required computation time for different implementations is listed when the number of data flows M is 100 and 200. The stopping criterion is that all the three performance metrics are below 10^{-3}, and the number of iterations is defined as the number of updates in the NC 0. The corresponding cumulative distribution function (CDF) of the computation time for CPU is also shown in Fig. 4.8. One can observe that after applying the parallel implementation across different nodes, the required computation time is much less than the commercial LP solver. Moreover, the Semi-BSUM-M algorithm can further reduce the computation time over the synchronous implementation, and the dynamic range of the computation time is much smaller than the synchronous counterpart. The variability of computation time is due to the fact that the size of each subnetwork is different to each other, so the computation time for local update differs. This demonstrates the benefits of semi-asynchronous model over the synchronous implementation. Furthermore, we should note that one iteration for synchronous implementation requires each of the 9 subnetworks exchange its latest information with NC 0. On the other hand, for the semi-asynchronous implementation, one central iteration only requires $|\mathcal{S}_{t,0}| \approx 1$ subnetwork exchanges their latest information with the NC 0. For fair comparison based on the communication overhead between nodes, on average 9 iterations for the asynchronous implementations account for 1 iteration for the synchronous implementation. Hence, the communication efficiency of the Semi-BSUM-M scheme is very close to that of the synchronous counterpart. A similar performance trend can also be observed for the number of data flows up to 200, showing the scalability of the proposed algorithms.

	M=100		M=200	
Approaches	Time	# of Iterations	Time	# of Iterations
Gurobi	4.66s	N/A	12.01s	N/A
Synchronous	8.25s	237.14	18.27s	266.04
Synchronous with parallelization	2.47s	237.14	5.67s	266.04
Semi-Asynchronous with parallelization	1.32s	2003.8	3.11s	2161.0

Table 4.3: Comparison of computation time used by different implementations for hierarchical network traffic problem.

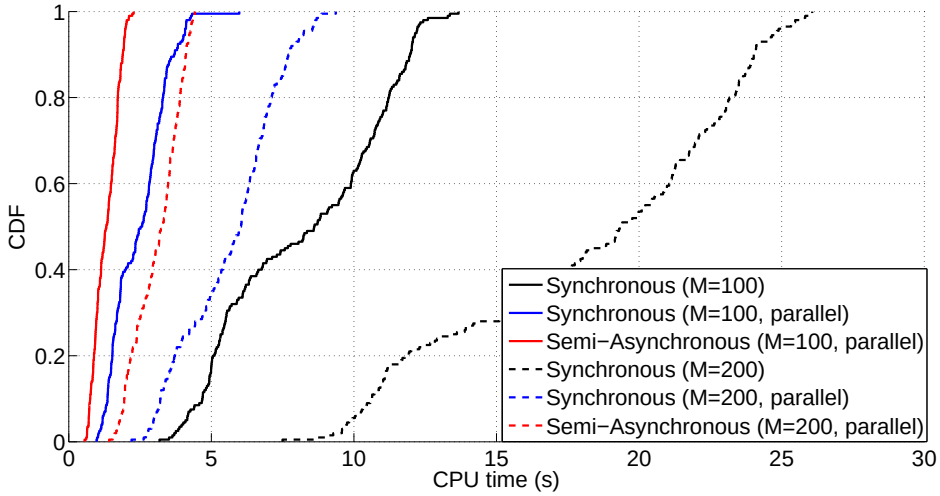


Figure 4.8: The CDF of the computation time used by different implementations for hierarchical network traffic problem.

Chapter 5

Conclusion and Discussion

In this dissertation, we consider i) the BS activation problem for HetNet; and ii) the network provisioning problem for C-RAN. For these design problems, we propose efficient and effective algorithms to respond the challenges arising from practical system limitations.

Specifically, in Chapter 2, We have utilized the sparsity-promoting techniques and proposed formulations and distributed algorithms that effectively select the active BSs. In Chapter 3, for the joint design problem of a C-RAN, the resources in both the fixed backhaul links and the wireless radio access links are optimized. Our proposed algorithm is capable of efficiently computing a high-quality solution for this large-scale and nonconvex problem in a distributed manner. However, the efficiency of the distributed implementations, e.g., algorithms in Chapter 2 and 3, highly depends on the synchronization requirement between the computation nodes. In Chapter 4, a semi-asynchronous distributed implementation is proposed. This implementation can well handle the asynchrony arising among the networked computation nodes, and its efficacy is demonstrated via solving the backhaul flow control problem of the C-RAN.

However, there are some practical issues for C-RAN that the dissertation has not addressed. In particular, we require the knowledge of the perfect CSI for all the wireless links, and no channel uncertainty is considered. Moreover, the capacities of the backhaul links should be known and fixed, so the effects of the imperfect lossy backhaul links [132] and the finite buffer size of the network routers need further investigation. For the manageability of the backhaul flow control, we further assume the BSs transmit signals

to users independently or with pre-determined BS clusters without any dynamic BS cooperation scheme. Furthermore, only the downlink transmission direction, i.e., from the BSs to the mobile users, is discussed in this dissertation. Although by changing the roles of the BSs and the mobile users, the proposed C-RAN algorithm can also handle the uplink data transmission, the coexistence of both directions is not a trivial extension for the practical TDMA system. With these observations, we identify the following important research directions for the provision of the future C-RAN architecture.

(i) **BS-user association with finite backhaul bandwidth:** Since the early works proposed by Yates [133] and Hanley [134], the optimal joint power allocation for system throughput and BS-user association has been a research focus for congestion control, traffic offloading, and avoiding the hot spots. Recently, this line of work has been extended to accommodate the multi-antenna transceiver [135, 136], and min rate utility function [137, 138]. However, these works do not consider the existence of multiple frequency tones and the finite bandwidth in the backhaul network. The proposed algorithm for C-RAN can dynamically choose the pathes for each commodity. Thus, it should be possible to properly address the issue of the BS-user association under theses extra constraints. Retrospectively, the association is crucial for the C-RAN algorithms since it reduces the complexity for managing the wireless links.

(ii) **Joint processing among BSs:** With multiple BSs, it is well-known that joint processing between BSs can greatly increase the achievable rates of mobile users. However, due to the finite capacity constraints on the backhaul network, sharing all users' signals among BSs may not be possible. It is interesting to extend our approach to the scenario that signals for users can be split into two separate parts, common and private information. The common information is shared among BSs for joint processing, and private information is only processed within each BS. Some initial results on this direction have been reported in [77] for the MISO downlink scenario, but the flow routing constraint in the backhaul network has not been considered. The backhaul throughput for multicast may be further increased by, for example, applying network coding [139]. This is also an important topic of future research.

(iii) **Dynamic BS clustering:** Despite the benefits of the joint processing among BSs, the management of the flows for commodities becomes much more complicated when the number of cooperating BSs increases. Therefore, extending the prior works on BS clustering for the cellular environment to the framework of C-RAN is needed. This direction is highly related to the subnetwork clustering such that the computation loading and communication overhead between NCs can be balanced, e.g., [140] [141].

(iv) **Coexistence of uplink and downlink information flows:** In TDMA systems, uplink and downlink information flows can occur on the same frequency. Hence, we are interested in extending our approach for the C-RAN architecture to this practical protocol. Specifically, additional sets of variables that control the airtime allocation between the uplink and downlink directions should be included, and the scheduling policy should be properly designed. These policies should also account for the transceiver hardware limitations. For example, the number of concurrent transmission beams at each wireless nodes cannot exceed the number of the RF chains. Similarly, at the same frequency, the two transmission directions cannot be activated at the same time, i.e., the half-duplex constraint. These extra limitations are summarized in [142], and their impacts on the traffic engineering problem should be investigated. Moreover, the admission control for both downlink and uplink users should also be considered when QoS requirements are imposed on mobile users. Recently, the BS activation problem with total power minimization design has been extended to the case that incorporates the uplink transmission [143].

(v) **Reducing the CSI overhead for wireless channels:** In the practical system design, the perfect CSI may not be available due to insufficient training frames or contaminated training signals. The large number of direct or indirect wireless links between BSs and users also prevent the system operator from obtaining the CSI for all links. Hence, it is crucial to extend our approach in C-RAN to accommodate this factor. To this end, one can assume only the availability of the long-term channel statistics, and consider the corresponding transceiver design problem [36] and [144]. Another possible approach is to extend the recent work on the utility

maximization with probabilistic rate constraints for the cellular environment [145] to the more general C-RAN architecture. Furthermore, it will be interesting to design resource allocation algorithms that can quickly adapt to account for the environment changes such as the coming and leaving of users.

References

[1] D. Gesbert, S. Hanly, H. Huang, S. Shamai (Shitz), O. Simeone, and W. Yu. Multi-cell MIMO cooperative networks: A new look at interference. *IEEE Journal on Selected Areas in Communications*, 28(9):1380–1408, Dec. 2010.

[2] A. Damnjanovic, J. Montojo, Y. Wei, T. Ji, T. Luo, M. Vajapeyam, T. Yoo, O. Song, and D. Malladi. A survey on 3GPP heterogeneous networks. *IEEE Trans. Wireless Communications*, 18(3):10–21, Jun. 2011.

[3] M. Hong and Z.-Q. Luo. Signal processing and optimal resource allocation for the interference channel. *Academic Press Library in Signal Processing: Volume 2 - Communications and Radar Signal Processing*, 2:409–469, 2014. available at http://arxiv.org/abs/1206.5144v1.

[4] E. Bjornson and E. Jorswieck. Optimal resource allocation in coordinated multi-cell systems. *Foundations and Trends in Communications and Information Theory*, 9(2-3):113–381, 2013.

[5] O. Blume, H. Eckhardtm, S. Klein, E. Kuehn, and W. M. Wajda. Energy savings in mobile networks based on adaptation to traffic statistic. *Bell Labs Technical Journal*, 15(2):77–94, Sep. 2010.

[6] O. Arnold, F. Richter, G. Fettweis, and O. Blume. Power consumption modeling of different base station types in heterogeneous cellular networks. In *Proc. of 2010 Future Network and Mobile Summit*, pages 1–8, Florence, Jun. 2010.

[7] G. Auer, V. Giannini, C. Desset, I. Godor, P. Skillermark, M. Olsson, M. Imran, D. Sabella, M. Gonzalez, O. Blume, , and A. Fehske. How much energy is needed

to run a wireless network? *IEEE Trans. Wireless Communications*, 18(5):40–49, Oct. 2011.

[8] J. Kani and H. Nakamura. Recent progress and continuing challenges in optical access network technologies. In *Proc. of the 2012 IEEE 3rd International Conference on Photonics (ICP)*, pages 66–70, Penang, Oct. 2012.

[9] A. Dhaini, P.-H. Ho, G. Shen, and B. Shihada. Energy efficiency in TDMA-based next-generation passive optical access networks. *IEEE/ACM Trans. Networking*, 22(3):850–863, Jun. 2014.

[10] J. Andrews. Seven ways that Hetnets are a cellular paradigm shift. *IEEE Communications Magazine*, 51(3):136–144, Mar. 2013.

[11] Huawei Technologies Inc. 5G: A technology vision, 2013. White paper.

[12] Open Networking Foundation. Software-defined networking: The new norm for networks, 2012. White paper.

[13] China Mobile. C-RAN: the road towards green RAN, Oct. 2011. White Paper, ver. 2.5.

[14] C.-L. I, J. Huang, R. Duan, C. Cui, X. Jiang, and L. Li. Recent progress on C-RAN centralization and cloudification. *IEEE Access*, 2:1030–1039, 2014.

[15] B. Clerckx, A. Lozano, S. Sesia, C. van Rensburg, and C. B. Papadias. 3GPP LTE and LTE Advanced. *EURASIP Journal on Wireless Communications and Networking*, 2009.

[16] J. Zhang, R. Chen, J.G. Andrews, A. Ghosh, and R.W. Heath. Networked MIMO with clustered linear precoding. *IEEE Trans. Wireless Communications*, 4(8):1910–921, Jun. 2009.

[17] A. Papadogiannis, D. Gesbert, and E. Hardouin. A dynamic clustering approach in wireless networks with multi-cell cooperative processing. In *Proc. of the 2008 IEEE International Conference on Communications*, pages 4033–4037, Beijing, May 2008.

[18] A. Papadogiannis and G.C. Alexandropoulos. The value of dynamic clustering of base stations for future wireless networks. In *2010 IEEE International Conference on Fuzzy Systems (FUZZ)*, pages 1–6, Barcelona, Jul. 2010.

[19] Y. Zeng, E. Gunawan, Y.-L. Guan, and J. Liu. Joint base station selection and linear precoding for cellular networks with multi-cell processing. In *TENCON*, pages 1976–1981, Fukuoka, Nov. 2010.

[20] S.-J. Kim, S. Jain, and G. B. Giannakis. Backhaul-constrained multi-cell cooperation using compressive sensing and spectral clustering. In *Proc. of the 2012 IEEE 13th Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pages 65–69, Cesme, Jun. 2012.

[21] M. Hong, R. Sun, and Z.-Q. Luo. Joint base station clustering and beamformer design for partial coordinated transmission in heterogenous networks. *IEEE Journal on Selected Areas in Communications*, 31(2):226–240, Feb 2013.

[22] Y. Cheng, M. Pesavento, and A. Philipp. Joint network optimization and downlink beamforming for CoMP transmissions using mixed integer conic programming. *IEEE Trans. Signal Process.*, 61(16):3972–3987, Aug. 2013.

[23] J. Zhao, T. Q. S. Quek, and Z. Lei. Coordinated multipoint transmission with limited backhaul data transfer. *IEEE Trans. Wireless Communications*, 12(6):2762–2775, Jun. 2013.

[24] B. Dai and W. Yu. Sparse beamforming for limited-backhaul network MIMO system via reweighted power minimization. In *Proc. of 2013 Global Communications Conference (GLOBECOM)*, pages 1962–1967, Atlanta, GA, Dec. 2013.

[25] B. Dai and W. Yu. Sparse beamforming and user-centric clustering for downlink cloud radio access network. *IEEE Access*, 2:132–1339, 2014.

[26] M. Bengtsson and B. Ottersten. Optimal and suboptimal transmit beamforming. Chapter 18 in *Handbook of Antennas in Wireless Communications*, L. C. Godara, Ed., CRC Press, Aug. 2001.

[27] A. Wiesel, Y. C. Eldar, and S. Shamai (Shitz). Linear precoding via conic optimization for fixed MIMO receivers. *IEEE Trans. Signal Processing*, 54(1):161–176, Jan. 2006.

[28] W. Yu and T. Lan. Transmitter optimization for the multi-antenna downlink with per-antenna power constraints. *IEEE Trans. Signal Process.*, 55(6):2646–2660, Jun. 2007.

[29] H. Dahrouj and W. Yu. Coordinated beamforming for the multicell multi-antenna wireless system. *IEEE Trans. Wireless Communications*, 9(5):1748–1759, May 2010.

[30] V. R. Cadambe and S. A. Jafar. Interference alignment and degrees of freedom of the k-user interference channel. *IEEE Trans. Information Theory*, 54(8):3425–3441, Aug. 2008.

[31] S. Mohajer, S. N. Diggavi, C. Fragouli, and D. Tse. Transmission techniques for relay-interference networks. In *Proc. of the 2008 46th Annu. Allerton Conf.*, pages 467–474, Urbana-Champaign, IL, Sep. 2008.

[32] S. Mohajer, D. Tse, and S. Diggavi. Approximate capacity of a class of gaussian relayinterference networks. In *Proc. of the 2009 IEEE International Symposium on Information Theory (ISIT)*, pages 31–35, Seoul, June 28-July 3 2009.

[33] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[34] M. Grant and S. Boyd. CVX: Matlab software for disciplined convex programming. http://stanford.edu/∼boyd/cvx, June 2009.

[35] F. Rashid-Farrokhi, K. J. R. Liu, and L. Tassiulas. Transmit beamforming and power control for cellular wireless systems. *IEEE Journal on Selected Areas in Communications*, 16(8):1437–1450, Oct. 1998.

[36] C. Shen, T.-H. Chang, K.-Y. Wang, Z. Qiu, and C.-Y. Chi. Distributed robust multicell coordinated beamforming with imperfect CSI: An ADMM approach. *IEEE Trans. Signal Processing*, 60(6):2988–3003, Jun. 2012.

[37] S. Joshi, M. Codreanu, and M. Latva-aho. Distributed resource allocation for MISO downlink systems via the alternating direction method of multipliers. *EURASIP Journal on Wireless Communications and Networking*, (1):1–19, Jan. 2014.

[38] D. P. Bertsekas and J. N. Tsitsiklis. *Parallel and Distributed Computation: Numerical Methods*. Athena Scientific, 1997.

[39] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122, 2011.

[40] M. Yuan and Y. Lin. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B*, 68(1):49–67, Feb. 2006.

[41] E. Candes, M. B. Wakin, and S. Boyd. Enhancing sparsity by reweighted l_1 minimization. *Journal on Fourier Analysis and Applications*, 14(5):877–905, Dec. 2008.

[42] Z.-Q. Luo and S. Zhang. Dynamic spectrum management: Complexity and duality. *IEEE Journal on Selected Topics in Signal Process.*, 2(1):57–73, Feb. 2008.

[43] W. Yu, G. Ginis, and J. Cioffi. Distributed multiuser power control for digital subscriber lines. *IEEE Journal on Selected Areas in Communications*, 20(5):1105–1115, Jun. 2002.

[44] R. Cendrillon, J. Huang, M. Chiang, and M. Moonen. Autonomous spectrum balancing for digital subscriber lines. *IEEE Trans. Signal Processing*, 55(8):4241–4257, Aug. 2007.

[45] R. Cendrillon, W. Yu, M. Moonen, J. Verlinden, and T. Bostoen. Optimal multiuser spectrum balancing for digital subscriber lines. *IEEE Trans. Communications*, 54(5):922–933, May 2006.

[46] J. Papandriopoulos and J. S. Evans. SCALE: A low-complexity distributed protocol for spectrum balancing in multiuser DSL networks. *IEEE Trans. Information Theory*, 55(8):3711–3724, Aug. 2009.

[47] S. Chung, S. J. Kim, J. Lee, and J. Cioffi. A game-theoretic approach to power allocation in frequency-selective gaussian interference channels. In *Proc. of the 2003 IEEE International Symposium on Information Theory (ISIT)*, page 316, June 29- July 4 2003.

[48] Z.-Q. Luo and J.-S. Pang. Analysis of iterative waterfilling algorithm for multiuser power control in digital subscriber lines. *EURASIP Journal on Advanced Signal Processing*, 2009:1–10, Mar. 2006.

[49] M. Kobayashi and G. Caire. Iterative waterfilling for weighted rate sum maximization in MIMO-OFDM broadcast channels. In *Proc. of 2007 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5–8, Honolulu, HI, Apr. 2007.

[50] R. Etkin, A. Parekh, and D. Tse. Spectrum sharing for unlicensed bands. *IEEE Journal on Selected Areas in Communications*, 25(3):517–528, Apr. 2007.

[51] K. W. Shum, K. K. Leung, and C. W. Sung. Convergence of iterative waterfilling algorithm for gaussian interference channels. *IEEE Journal on Selected Areas in Communications*, 25(6):1091–1100, Aug. 2007.

[52] C. Shi, R. Berry, and M. Honig. Distributed interference pricing with MISO channels. In *Proc. of the 2008 46th Annu. Allerton Conf.*, pages 539–546, Urbana-Champaign, IL, Sep. 2008.

[53] C. Shi, R. Berry, and M. L. Honig. Monotonic convergence of distributed interference pricing in wireless networks. In *Proc. of the 2009 IEEE International Symposium on Information Theory (ISIT)*, pages 1619–1623, Seoul, June 28-July 3 2009.

[54] S. S. Christensen, R. Agarwal, E. Carvalho, and J. M. Cioffi. Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design. *IEEE Trans. Wireless Communications*, 7(12):4792–4799, Dec. 2008.

[55] D. A. Schmidt, C. Shi, R. A. Berry, M. L. Honig, and W. Utschick. Minimum mean squared error interference alignment. In *Asilomar Conference on Signals, Systems and Computers*, pages 1106–1110, Pacific Grove, CA, Nov. 2009.

[56] Q. Shi, M. Razaviyayn, Z.-Q. Luo, and C. He. An iteratively weighted mmse approach to distributed sum-utility maximization for a MIMO interfering broadcast channel. *IEEE Trans. on Signal Process.*, 59(9):4331–4340, Sep. 2011.

[57] P.-Q. Li, Y. A. Zhang, and J. Huang. MAPEL: Achieving global optimality for a non-convex wireless power control problem. *IEEE Trans. Wireless Communications*, 8(3):1553–1563, Mar. 2009.

[58] G. Scutari, F. Facchinei, P. Song, D. P. Palomar, and J.-S. Pang. Decomposition by partial linearization : Parallel optimization of multi-agent systems. *IEEE Trans. Signal Process.*, 62(3):641–656, Feb 2014.

[59] D. A. Schmidt, C. Shi, R. A. Berry, M. L. Honig, and W. Utschick. Comparison of distributed beamforming algorithms for MIMO interference networks. *IEEE Trans. Signal Process.*, 61(13):3476–3489, Jul. 2013.

[60] P. Hande, S. Rangan, M. Chiang, and X. Wu. Distributed uplink power control for optimal sir assignment in cellular data networks. *IEEE/ACM Trans. Networking*, 16(6):1420–1433, Dec. 2008.

[61] R. Zhang and S. Cui. Cooperative interference management with MISO beamforming. *IEEE Trans. Signal Process.*, 58(10):5450–5458, Oct. 2010.

[62] Y.-F Liu, M. Hong, and Y.-H. Dai. Max-min fairness linear transceiver design problem for a multi-user SIMO interference channel is polynomial time solvable. *IEEE Signal Processing Letters*, 20(1):27 –30, Jan. 2013.

[63] Y.-F Liu, Y.-H. Dai, and Z.-Q. Luo. Max-min fairness linear transceiver design for a multi-user MIMO interference channel. *IEEE Trans. Signal Process.*, 61(9):2413–2423, May 2013.

[64] M. Razaviyayn, M. Hong, and Z.-Q. Luo. Linear transceiver design for a MIMO interfering broadcast channel achieving max-min fairness. *Signal Processing*, 93(12):3327–3340, Dec. 2013.

[65] Y. Shi, J. Zhang, and K. B. Letaief. Group sparse beamforming for green cloud-RAN. *IEEE Trans. Wireless Communications*, 13(5):2809–2823, May 2014.

[66] J. Li, E. Bjornson, T. Svensson, T. Eriksson, and M. Debbah. Joint precoding and load balancing optimization for energy-efficient heterogeneous networks. *IEEE Trans. Wireless Communications*, 14(10):5810–5822, Oct. 2015.

[67] B. R. Marks and G. P. Wright. A general inner approximation algorithm for nonconvex mathematical programs. *Operations Research*, 26(4):681–683, 1978.

[68] M. Razaviyayn, M. Hong, and Z.-Q. Luo. A unified convergence analysis of block successive minimization methods for nonsmooth optimization. *SIAM Journal on Optimization*, 23(2):1126–1153, 2013.

[69] D. P. Bertsekas. *Network Optimization: Continuous and Discrete Models*. Athena Scientific, 1998.

[70] D. P. Bertsekas, P. Hosein, and P. Tseng. Relaxation methods for network flow problems with convex arc costs. *SIAM Journal on Control and Optimization*, 25(5):1219–1243, Sep. 1987.

[71] D. P. Bertsekas and P. Tseng. The relax codes for linear minimum cost network flow problem. *Annals of Operations Research*, 13(1):125–190, 1988.

[72] F. Kelly, A. K. Maulloo, and D. Tan. Rate control for communication networks: Shadow prices, proportional fairness and stability. *The Journal of the Operational Research Society*, 49(3):237–252, 1998.

[73] F. Kelly. Global optimization of network flows. *preprint*, 2014.

[74] W. Yu, T. Kwon, and C. Shin. Multicell coordination via joint scheduling, beamforming, and power spectrum adaptation. In *Proc. of the 2011 IEEE INFOCOM*, pages 2570–2578, Shanghai, Apr. 2011.

[75] A. Chowdhery, W. Yu, and J. M. Cioffi. Cooperative wireless multicell OFDMA network with backhaul capacity constraints. In *Proc. of the 2011 IEEE International Conference on Communications*, pages 1–6, Kyoto, Jun. 2011.

[76] S. Mehryar, A. Chowdhery, and W. Yu. Dynamic cooperation link selection for network MIMO systems with limited backhaul capacity. In *Proc. of the 2012*

IEEE International Conference on Communications, pages 4410–4415, Ottawa, ON, Jun. 2012.

[77] R. Zakhour and D. Gesbert. Optimized data sharing in multicell MIMO with finite backhaul capacity. *IEEE Trans. Signal Process.*, 59(12):6102–6111, Dec. 2011.

[78] L. Xiao, M. Johansson, and S. P. Boyd. Simultaneous routing and resource allocation via dual decomposition. *IEEE Trans. Communications*, 52(7):1136–1144, Jul. 2004.

[79] M. J. Neely, E. Modiano, and C. E. Rohrs. Dynamic power allocation and routing for time-varying wireless networks. *IEEE Journal on Selelected Areas in Communication*, 23(1):89–103, Jan. 2005.

[80] J. Yuan, Z. Li, W. Yu, and B. Li. A cross-layer optimization framework for multihop multicast in wireless mesh networks. *IEEE Journal on Selected Areas in Communications*, 24(11):2092–2103, Nov. 2006.

[81] A. Ribeiro and G. B. Giannakis. Separation principles in wireless networking. *IEEE Trans. Information Theory*, 56(9):4488–4505, Sep. 2010.

[82] Z. Shao, M. Chen, A. S. Avestimehr, and S.-Y. Robert Li. Cross-layer optimization for wireless networks with deterministic channel models. *IEEE Trans. Information Theory*, 57(9):5840–5862, Sep. 2011.

[83] M. Mardani, S.-J. Kim, and G. B. Giannakis. Cross-layer design of wireless multihop random access networks. *IEEE Trans. Signal Process.*, 60(5):2562–2574, May 2012.

[84] E. Matskani, N. D. Sidiropoulos, and L. Tassiulas. Convex approximation algorithms for back-pressure power control. *IEEE Trans. Signal Process.*, 60(4):1957–1970, Apr. 2012.

[85] B. Gopalakrishnan and N. D. Sidiropoulos. Joint back-pressure power control and interference cancellation in wireless multi-hop networks. *IEEE Trans. Wireless Communications*, 12(7):3484–3495, Jul. 2013.

[86] L. Georgiadis, M. J. Neely, and L. Tassiulas. Resource allocation and cross-layer control in wireless networks. *Foundations and Trends in Networking*, 1(1):1–144, 2006.

[87] M. Chiang, S. Low, A. R. Calderbank, and J. C. Doyle. Layering as optimization decompositioon: A mathematical theory of network architectures. *Proceedings of the IEEE*, 95(1):255–312, Jan. 2007.

[88] N. Shroff and R. Srikant. A tutorial on cross-layer optimization in wireless networks. *IEEE Journal on Selected Areas in Communications*, 24(8):1452–1463, Aug. 2006.

[89] A. Avestimehr, S. Diggavi, and D. Tse. Wireless network information flow: A deterministic approach. *IEEE Trans. on Information Theory*, 57(4):5840–5862, Apr. 2011.

[90] M. Zheng, S. Stanczak, and H. Yu. Utility-cost optimization for joint routing and power control in multi-hop wireless networks. In *Proc. of the 2012 IEEE Wireless Communications and Networking Conference (WCNC)*, pages 1665–1669, Shanghai, Apr. 2012.

[91] D. P. Bertsekas. *Nonlinear Programming.* Athena Scientific, 1995.

[92] S. H. Low and D. E. Lapsley. Optimization flow control, I: Basic algorithm and convergence. *IEEE/ACM Trans. Networking*, 7(6):861–874, Dec. 1999.

[93] A. Ghosh, S. Ha, E. Crabbe, and J. Rexford. Scalable multi-class traffic management in data center backbone networks. *IEEE Journal on Selected Areas in Communications*, 31(12):2673–2684, Dec. 2013.

[94] M. Leinonen, M. Codreanu, and M. Juntti. Distributed joint resource and routing optimization in wireless sensor networks via alternating direction method of multipliers. *IEEE Trans. Wireless Communications*, 12(11):5454–5467, Nov. 2013.

[95] H. Xu and B. Li. Joint request mapping and response routing for geo-distributed cloud services. In *Proc. IEEE INFOCOM*, pages 854–862, Turin, Apr. 2013.

[96] S. Joshi, M. Codreanu, and M. Latva-aho. Distributed resource allocation for MISO downlink systems via the alternating direction method of multipliers. In *Asilomar Conference on Signals, Systems and Computers*, pages 488–493, Pacific Grove, CA, Nov. 2012.

[97] C. Feng, H. Xu, and B. Li. An alternating direction method approach to cloud traffic management. *submitted to IEEE/ACM Trans. Networking*, 2014.

[98] Y. Shi, J. Zhang, K. B. Letaief, B. Bai, and W. Chen. Large-scale convex optimization for ultra-dense cloud-RAN. *IEEE Trans. Wireless Magazine*, 22(3):84–91, Jun. 2015.

[99] Y. Low, D. Bickson, J. Gonzalez, C. Guestrin, A. Kyrola, and J. M. Hellerstein. Distributed GraphLab: a framework for machine learning and data mining in the cloud. *Proc. of the VLDB Endowment*, 5(8):716–727, Apr. 2012.

[100] L. Bui, A. Eryilmaz, R. Srikant, and W. Xinzhou. Asynchronous congestion control in multi-hop wireless networks with maximal matching-based scheduling. *IEEE/ACM Trans. Networking*, 16(4):826–839, Aug. 2008.

[101] O. Simeone, O. Somekh, H. Poor, and S. Shamai (Shitz). Downlink multicell processing with limited-backhaul capacity. *EURASIP Journal on Advanced Signal Processing*, 2009:1–10, Feb. 2009.

[102] S.-H. Park, O. Simeone, O. Sahin, and S. Shamai (Shitz). Joint precoding and multivariate backhaul compression for the downlink of cloud radio access networks. *IEEE Trans. Signal Process.*, 61(22):5646–5658, Nov. 2013.

[103] P. Patil, B. Dai, and Wei Yu. Performance comparison of data-sharing and compression strategies for cloud radio access networks. In *Proc. of the 2015 European Signal Processing Conference (EUSIPCO)*, Nice, France, Aug. 31-Sep. 4 2015.

[104] S.-H. Park, O. Simeone, O. Sahin, and S. Shamai (Shitz). Multihop backhaul compression for the uplink of cloud radio access networks. *IEEE Trans. Vehicular Technology*, 2015.

[105] O. Mehanna, N. D. Sidiropoulos, and G. B. Giannakis. Multicast beamforming with antenna selection. In *Proc. of the 2012 IEEE 13th Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pages 70–74, Cesme, Jun. 2012.

[106] Y.-F. Liu, Y.-H. Dai, and Z.-Q. Luo. Joint power and admission control via linear programming deflation. *IEEE Trans. Signal Process.*, 61(6):1327–1338, Mar. 2013.

[107] S. Jain, S.-J. Kim, and G. Giannakis. Backhaul-constrained multi-cell cooperation leveraging sparsity and spectral clustering. *to appear in IEEE Trans. Wireless Communications*, Jun. 2015.

[108] M. Hong, T.-H. Chang, X. Wang, M. Razaviyayn, S. Ma, and Z.-Q. Luo. A block successive upper bound minimization method of multipliers for linearly constrained convex optimization. *submitted for publication, available at http://arxiv.org/abs/1401.7079*, 2014.

[109] R. Zhang and J. T. Kwok. Asynchronous distributed ADMM for consensus optimization. In *Proc. of the 31st International Conference on Machine Learning (ICML-14)*, pages 1701–1709, Beijing, Jun. 2014.

[110] M. Hong. A distributed, asynchronous and incremental algorithm for nonconvex optimization: An admm based approach. *submitted for publication, available at http://arxiv.org/abs/1412.6058v1*, 2014.

[111] E. Wei and A. Ozdaglar. On the O(1/k) convergence of asynchronous distributed alternating direction method of multipliers. In *Proc. of 2013 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pages 551–554, Austin, TX, Dec. 2013.

[112] Z. Peng, Y. Xu, M. Yan, and W. Yin. Arock: an algorithmic framework for asynchronous parallel coordinate updates. *UCLA CAM Report*, pages 15–37, 2015.

[113] M. Hong, Z.-Q. Luo, and M. Razaviyayn. Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems. *submitted for publication, available at http://arxiv.org/abs/1410.1390*, 2014.

[114] S. J. Wright, R. D. Nowak, and M. A. T. Figueiredo. Sparse reconstruction by separable approximation. *IEEE Trans. Signal Process.*, 57(7):2479–2493, Jul. 2009.

[115] H. Du, T. Ratnarajah, M. Pesavento, and C. B. Papadias. Joint transceiver beamforming in MIMO cognitive radio network via second-order cone programming. *IEEE Trans. Signal Process.*, 60(2):781–792, Feb. 2012.

[116] M. Hong, M. Razaviyayn, R. Sun, and Z.-Q. Luo. Joint transceiver design and base station clustering for heterogeneous networks. In *Asilomar Conference on Signals, Systems and Computers*, pages 574 – 578, Pacific Grove, CA, Nov. 2012.

[117] P. Tseng. Convergence of a block coordinate descent method for nondifferentiable minimization. *Journal of Optimization Theory and Applications*, 109(3):475–494, Jun. 2001.

[118] C.-X. Wang, X. Hong, X. Ge, X. Cheng, G. Zhang, and J. Thompson. Cooperative MIMO channel models: A survey. *IEEE Communications Magazine*, 48(2):80–87, Feb. 2010.

[119] E. Danna, S. Mandal, and A. Singh. A practical algorithm for balancing the max-min fairness and throughput objectives in traffic engineering. In *Proc. of the 2012 IEEE INFOCOM*, pages 846–854, Orlando, FL, Mar. 2012.

[120] B. Dai and W. Yu. Backhaul-aware multicell beamforming for downlink cloud radio-access network. In *Proc. of IEEE International Conference on Communications (ICC) Workshop on Cloud-Processing in Heterogeneous Mobile Communication Networks*, London, Jun. 2015.

[121] Gurobi Optimization, Inc. Gurobi optimizer reference manual, 2013.

[122] F. Niu, B. Recht, C. Re, and S. J. Wright. Hogwild!: A lock-free approach to parallelizing stochastic gradient descent. In *Advances in Neural Information Processing Systems (NIPS)*, pages 693–701, Granada, Dec. 2011.

[123] A. Agarwal and J. C. Duchi. Distributed delayed stochastic optimization. In *Advances in Neural Information Processing Systems (NIPS)*, pages 873–881, Granada, Dec. 2011.

[124] J. Liu, C. Re, S. J. Wright, and V. Bittorf. An asynchronous parallel stochastic coordinate descent algorithm. In *Proc. of the 31st International Conference on Machine Learning (ICML-14)*, pages 1701–1709, Beijing, Jun. 2014.

[125] M. Li, D. G. Andersen, and A. Smola. Distributed delayed proximal gradient methods. In *NIPS Workshop on Optimization for Machine Learning*, Lake Tahoe, NV, Dec. 2013.

[126] B. He, M. Tao, and X. Yuan. Alternating direction method with Gaussian back substitution for separable convex programming. *SIAM Journal on Optimization*, 22(2):313–340, 2012.

[127] B. He, H. Xu, and X. Yuan. On the proximal Jacobian decomposition of ALM for multiple-block separable convex minimization problems and its relationship to ADMM. *Journal of Scientific Computing*, Jun. 2015.

[128] W. Deng, M.-J. Lai, Z. Peng, and W. Yin. Parallel multi-block ADMM with o(1/k) convergence. *UCLA CAM Report*, pages 13–64, 2013.

[129] J.-S. Liu and C.-H. Lin. Asynchronous distributed joint optimization in wireless multi-hop networks. *IEEE Communications Letters*, 2015.

[130] C. Chen, B. He, Y. Ye, and X. Yuan. The direct extension of admm for multi-block convex minimization problems is not necessarily convergent. *Mathematical Programming*, 2014.

[131] X. Wang, M. Hong, S. Ma, and Z.-Q. Luo. Solving multiple-block separable convex minimization problems using two-block alternating direction method of multipliers. *Pacific Journal of Optimization*, 2015.

[132] Q. Gao, J. Zhang, and S. Hanly. Cross-layer rate control in wireless networks with lossy links: Leaky-pipe flow, effective network utility maximization and hop-by-hop algorithms. *IEEE Trans. Wireless Communications*, 8(6):3068–3076, Jun. 2009.

[133] R. D. Yates. A framework for uplink power control in cellular radio systems. *IEEE Journal on Selected Areas in Communications*, 13(7):1341–1347, Sep. 1995.

[134] S. V. Hanly. An algorithm for combined cell-site selection and power control to maximize cellular spread spectrum capacity. *IEEE Journal on Selected Areas in Communications*, 13(7):1332–1340, Sep. 1995.

[135] K. Shen and W. Yu. Downlink cell association optimization for heterogeneous networks via dual coordinate descent. In *Proc. of 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4779–4783, Vancouver, BC, May 2013.

[136] S. Sardellitti, G. Scutari, and S. Barbarossa. Joint cell selection aand radio resource allocation in MIMO small cell networks via successive convex approximation. In *Proc. of 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 850–854, Florence, May 2014.

[137] R. Sun, M. Hong, and Z.-Q. Luo. Optimal joint base station assignment and power allocation in a cellular network. *Proc. of the 2012 IEEE 13th Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pages 234–238, Jun. 2012.

[138] R. Sun and Z.-Q. Luo. Globally optimal joint uplink base station association and power control for max-min fairness. In *Proc. of 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 454–458, Florence, May 2014.

[139] Z. Li, B. Li, D. Jiang, and L. C. Lau. On achieving optimal throughput with network coding. In *Proc. of the 2005 IEEE INFOCOM*, pages 2184–2194, Mar. 2005.

[140] X. Li, P. Djukic, and H. Zhang. Zoning for hierarchical network optimization in software defined networks. In *Proc. of 1st IEEE International Workshop on SDN Management and Orchestration (SDNMO2014)*, pages 1–8, Krakow, May 2014.

[141] X. Li and H. Zhang. Creating logical zones for hierarchical traffic engineering optimization in SDN-empowered 5G. In *Proc. of the 2015 International Conference on Computing Networking and Communications (ICNC)*, pages 1071–1075, Garden Grove, CA, Feb. 2015.

[142] A. Callard and A. Stephenne. High frequency backhaul design and configuration. *in manuscript*, 2015.

[143] S. Luo, R. Zhang, and T. J. Lim. Downlink and uplink energy minimization through user association and beamforming in C-RAN. *IEEE Trans. Wireless Communications*, 14(1):494–508, Jan. 2015.

[144] M. Razaviyayn, M. Sanjabi, and Z.-Q. Luo. A stochastic weighted MMSE approach to sum rate maximization for a MIMO interference channel. *Proc. of the 2013 IEEE 14th Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pages 325–329, Jun. 2013.

[145] W.-C. Li, T.-H. Chang, C. Lin, and C.-Y. Chi. Coordinated beamforming for multisuer MISO interference channel under rate outage constraints. *IEEE Trans. Signal Processing*, 61(5):1087–1103, Mar. 2013.

[146] H. Robbins and D. Siegmund. A convergence theorem for non-negative almost supermartingales and some applications. *Optimizing Methods in Statistics. Academic Press, New York*, pages 233–257, 1971.

Appendix A

A Brief Review of the ADMM Algorithm

The ADMM algorithm was originally developed in 1970s, and has attracted lots of interests recently due to its efficiency in large-scale optimization (see [39] and references therein). Specifically, the ADMM is designed to solve the following structured convex problem

$$\begin{aligned} \min_{\mathbf{x}\in\mathbb{C}^n,\mathbf{z}\in\mathbb{C}^m} \quad & f(\mathbf{x}) + g(\mathbf{z}) \\ \text{s.t.} \quad & \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z} = \mathbf{c} \\ & \mathbf{x} \in \mathcal{C}_1,\ \mathbf{z} \in \mathcal{C}_2 \end{aligned} \tag{A.1}$$

where $\mathbf{A} \in \mathbb{C}^{k\times n}$, $\mathbf{B} \in \mathbb{C}^{k\times m}$, $\mathbf{c} \in \mathbb{C}^k$, and f and g are convex functions while $\mathcal{C}_1$ and $\mathcal{C}_2$ are non-empty convex sets.

The partial augmented Lagrangian function for problem (A.1) can be expressed as

$$L_\rho(\mathbf{x},\mathbf{z},\mathbf{y}) = f(\mathbf{x}) + g(\mathbf{z}) + \mathrm{Re}\left(\mathbf{y}^H(\mathbf{A}\mathbf{x}+\mathbf{B}\mathbf{z}-\mathbf{c})\right) + (\rho/2)\|\mathbf{A}\mathbf{x}+\mathbf{B}\mathbf{z}-\mathbf{c}\|_2^2 \tag{A.2}$$

where $\mathbf{y} \in \mathbb{C}^k$ is the Lagrangian dual variables associated with the linear equality constraint, and $\rho > 0$ is some constant. The ADMM algorithm solves problem (A.1) by

iteratively performing three steps in each iteration t:

$$\mathbf{x}^{(t)} = \arg\min_{\mathbf{x}} L_\rho(\mathbf{x}, \mathbf{z}^{(t-1)}, \mathbf{y}^{(t-1)}) \tag{A.3a}$$

$$\mathbf{z}^{(t)} = \arg\min_{\mathbf{z}} L_\rho(\mathbf{x}^{(t)}, \mathbf{z}, \mathbf{y}^{(t-1)}) \tag{A.3b}$$

$$\mathbf{y}^{(t)} = \mathbf{y}^{(t-1)} + \rho(\mathbf{A}\mathbf{x}^{(t)} + \mathbf{B}\mathbf{z}^{(t)} - \mathbf{c}). \tag{A.3c}$$

In many applications, the primal subproblems (A.3a) and (A.3b) can be solved easily in closed-form, leading to the efficiency of ADMM. The convergence property of this algorithm is summarized in the following proposition [38, Proposition 4.2].

Proposition 1 *Assume that the optimal solution set of problem* (A.1) *is non-empty, and* $\mathbf{A}^T\mathbf{A}$ *and* $\mathbf{B}^T\mathbf{B}$ *are invertible. Then the sequence of* $\{\mathbf{x}^{(t)}, \mathbf{z}^{(t)}, \mathbf{y}^{(t)}\}$ *generated by* (A.3a), (A.3b), *and* (A.3c) *is bounded and every limit point of* $\{\mathbf{x}^{(t)}, \mathbf{z}^{(t)}\}$ *is an optimal solution of problem* (A.1).

Appendix B

Proof of Theorem 1

To prove Theorem 1, it is sufficient to show that problem (2.6) is strongly NP-hard. Consider a simple single-cell network with Q single antenna BSs serving Q users. That is, $K = 1$, $N = 1$, $|\mathcal{B}_k| = |\mathcal{U}_k| = Q$. Then problem (2.6) can be simplified to

$$\begin{aligned} \min_{\{p_i^q\}} \ & \sum_{i=1}^{Q} \left\| \sum_{q=1}^{Q} p_i^q \right\|_0 \\ \text{s.t.} \ & \frac{\sum_{q=1}^{Q} p_i^q g_i^q}{\sigma_i^2 + \sum_{j \neq i} \sum_{q=1}^{Q} p_j^q g_i^q} \geq \tau_i, \\ & \sum_{i=1}^{Q} p_i^q \leq P_q, \ p_i^q \geq 0, \quad \forall \ i, q = 1, \ldots, Q, \end{aligned} \tag{B.1}$$

where we have omitted the cell index k, and have defined $p_i^q \triangleq \|\mathbf{v}_i^q\|_2^2$ and $g_i^q \triangleq \|\mathbf{h}_i^q\|_2^2$, $\forall i, q = 1, \ldots, Q$. We prove that problem (B.1) is strongly NP-hard by establishing a polynomial time transformation from the so-called vertex cover problem. The vertex cover problem can be described as follows: given a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and a positive integer $W \leq |\mathcal{V}|$, we are asked whether there exists a vertex cover of size W or less, i.e., a subset $\mathcal{V}' \subset \mathcal{V}$ such that $|\mathcal{V}'| \leq W$, and for each edge $\{u, v\} \in \mathcal{E}$ at least one of the end points u and v belongs to $\mathcal{V}$.

Given a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with $|\mathcal{V}| = Q$, we let

$$g_i^q = g_q^i = \begin{cases} 1, & \text{if } i = q \text{ or } (i,q) \in \mathcal{E} \\ 0, & \text{if } (i,q) \notin \mathcal{E} \end{cases}$$

$$\tau_i = \frac{1}{Q^2},\ \sigma_i^2 = Q,\ P_q = Q,\ \ \forall q = 1, \ldots, Q.$$

We claim that the optimal value of problem (B.1) is less than or equal to W if and only if there exists a vertex cover set $\mathcal{V}'$ for the graph satisfying $|\mathcal{V}'| \leq W$.

"If" direction: Let $\mathcal{V}'$ with $|\mathcal{V}'| \leq W$ be the vertex cover set for the graph $\mathcal{G}$. Without loss of generality, suppose $\mathcal{V}' = \{1, 2, \ldots, W\}$. Then we can construct a feasible solution for problem (B.1) based on the cover set $\mathcal{V}'$ such that the optimal value of problem (B.1) at this point is equal to W. In particular, we have

$$p_i^q = 1,\ i = 1, \ldots, Q,\ q = 1, 2, \ldots, W$$
$$p_i^q = 0,\ i = 1, \ldots, Q,\ q = W+1, W+2, \ldots, Q$$

Therefore, the choice of $\{p_i^q\}_{i,q}$ satisfies $\sum_{i=1}^{Q} p_i^q = W \leq Q,\ \forall\ i, q = 1, \ldots, Q$, and the nonnegative constraints. Next, we check the feasibility of the above constructed solution.

- For user $i = 1, 2, \ldots, W$, the SINR constraint in (B.1) is satisfied, since $p_q^q = g_q^q = 1$ for all $q = 1, \ldots, W$. In particular, the satisfaction of the SINR constraint of user i, $\forall\ i$, can be derived as follows,

$$\frac{\sum_{q=1}^{Q} p_i g_i^q}{\sigma_i^2 + \sum_{j \neq i} \sum_{q=1}^{Q} p_j^q g_i^q} = \frac{\sum_{q=1}^{W} g_i^q}{Q + \sum_{j \neq i} \sum_{q=1}^{W} g_i^q} \geq \frac{1}{Q + (Q-1)W} \geq \frac{1}{Q^2} = \tau_i$$

- For user $i = W+1, W+2, \ldots, Q$, according to the definition of the cover set, there must exist $q \in \mathcal{V}'$ such that $(i,q) \in \mathcal{E}$ and thus $p_i^q = g_i^q = 1$. Hence, the SINR constraint of user $i = W+1, W+2, \ldots, Q$ are also satisfied.

"Only if" direction: Suppose that the optimal value of problem (B.1) is less than or equal to W and its optimal solution is $p_i^{q\star},\ \ \forall i, q = 1, \ldots, Q$. We construct the following sets

$$S_q \triangleq \{i \mid p_i^{q\star} g_i^q > 0\} = \{i \mid p_i^{q\star} > 0\},\ q = 1, \ldots, Q,$$

where the equality holds since when $g_i^q = 0$, $p_i^{q\star}$ should also be 0 to reduce the objective value while keeping the satisfaction of all the constraints. By the fact that the optimal value of problem (B.1) is less than or equal to W, we know that at most W of the defined sets S_q are nonempty sets. Without loss of generality, suppose these W nonempty sets are $S_1, \ldots, S_W$. Furthermore, the fact that all SINR constraints are satisfied, which ensures that $\sum_{q=1}^{Q} p_i^{q\star} g_i^q = \sum_{q=1}^{W} p_i^{q\star} g_i^q > 0$, $\forall\ i$. Combining this fact and the choice of $\{g_i^q\}$ based on the connectivity of node i and node q, we conclude that

$$\mathcal{V} = \bigcup_{q=1}^{Q} S_q = \bigcup_{q=1}^{W} S_q.$$

The above shows that $\{1, 2, \ldots, W\}$ constitutes a cover set of $\mathcal{V}$, which completes the proof. □

Appendix C

Proof of Theorem 4

In the following, $\{\mathbf{r}^\star, \mathbf{v}^\star; \boldsymbol{\delta}^\star, \boldsymbol{\theta}^\star, \kappa^\star, \boldsymbol{\epsilon}^\star\}$ is denoted as the KKT solutions of problem (3.5), which is restated in the following for reference,

$$\max_{\mathbf{v},\,\mathbf{r}} \quad r_{\min} \tag{C.1a}$$

$$\text{s.t.} \quad \mathbf{r} \geq 0,\ r_m \geq r_{\min},\ m = 1 \sim M, \tag{C.1b}$$

$$\mathbf{1}^T \mathbf{f}_l \leq C_l,\ \forall\, l \in \mathcal{L}^w, \tag{C.1c}$$

$$\mathbf{1}^T \mathbf{f}_l \leq \bar{f}_l(\mathbf{v}), \tag{C.1d}$$

$$\sum_{l \in \text{In}(v)} f_{l,m} + 1_{v=s_m} r_m = \sum_{l \in \text{Out}(v)} f_{l,m} + 1_{v=d_m} r_m,\ \ m = 1 \sim M,\ \forall\, v \in \mathcal{V}, \tag{C.1e}$$

$$\sum_{l \in \text{Out}(s) \bigcap \mathcal{L}^{wl}} v_l^2 \leq P_s,\ \forall\, s \in \mathcal{B}. \tag{C.1f}$$

Here $\boldsymbol{\delta}^\star, \boldsymbol{\theta}^\star, \boldsymbol{\kappa}^\star$, and $\boldsymbol{\epsilon}^\star$ respectively denote the corresponding optimal Lagrangian dual variables for the nonnegativeness constraints (C.1b) as well as {(C.1c), (C.1d)}, (C.1e), and (C.1f). The KKT solutions are similarly denoted as $\{\hat{\mathbf{r}}, \hat{\mathbf{v}}, \hat{\mathbf{u}}, \hat{\mathbf{w}}; \hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\kappa}}, \hat{\boldsymbol{\epsilon}}\}$ for problem (3.7), which is also restated as follows,

$$\max_{\mathbf{v},\,\mathbf{r},\,\mathbf{u},\,\boldsymbol{w}} \quad r_{\min} \tag{C.2a}$$

$$\text{s.t. (C.1b), (C.1c), (C.1e), and (C.1f)}, \tag{C.2b}$$

$$\mathbf{1}^T \mathbf{f}_l \leq 1 + \log(w_l) - w_l e_l(u_l, \mathbf{v}),\ \forall\, l \in \mathcal{L}^{wl}, \tag{C.2c}$$

where $\hat{\boldsymbol{\theta}}$ now is the Lagrangian dual variables for constraints {(C.1c), (C.2c)}.

Step 1: Denote $\boldsymbol{x}^\star \triangleq \{\mathbf{r}^\star, \mathbf{v}^\star; \boldsymbol{\delta}^\star, \boldsymbol{\theta}^\star, \boldsymbol{\kappa}^\star, \boldsymbol{\epsilon}^\star\}$ as an arbitrary KKT solution of problem (C.1)**. Then $\hat{\boldsymbol{y}} \triangleq \{\hat{\mathbf{r}}, \hat{\mathbf{v}}, \hat{\mathbf{u}}, \hat{\mathbf{w}}; \hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\kappa}}, \hat{\boldsymbol{\epsilon}}\} = \{\mathbf{r}^\star, \mathbf{v}^\star, \mathbf{u}(\mathbf{v}^\star), \mathbf{w}(\mathbf{v}^\star); \boldsymbol{\delta}^\star, \boldsymbol{\theta}^\star, \boldsymbol{\kappa}^\star, \boldsymbol{\epsilon}^\star\}$ is also a KKT solution of problem** (C.2)**, where we stack $u_l(\mathbf{v}^\star)$ and $w_l(\mathbf{v}^\star)$, $\forall\ l \in \mathcal{L}^{wl}$, as $\mathbf{u}(\mathbf{v}^\star)$ and $\mathbf{w}(\mathbf{v}^\star)$, respectively. The reverse statement is also true.**

Since some of the constraints of problem (C.1) and problem (C.2) are exactly the same, i.e., (C.1c), (C.1e), and (C.1f), the corresponding feasibility and the complementary slackness conditions of these constraints are of the same form for both problems. Hence, if $\boldsymbol{x}^\star$ can satisfy these constraints for problem (C.1), $\hat{\boldsymbol{y}}$ can satisfy those of problem (C.2). We only need to check the remaining KKT conditions given below. For problem (C.1), we have

$$-2\epsilon^\star_{s_l} v^\star_l + \sum_{n\in\bar{I}(l)} \theta^\star_n \nabla_{v_l} \bar{f}_n(\mathbf{v}^\star) = 0,\ \forall\, l \in \mathcal{L}^{wl}, \tag{C.3a}$$

$$-\delta^\star + \sum_{m=1}^{M} \delta^\star_m = 1, \tag{C.3b}$$

$$\delta^\star_m + \kappa^{s_m\star}_m - \kappa^{d_m\star}_m = 0,\ \forall\, m, \tag{C.3c}$$

$$\delta^\star_{l,m} - \theta^\star_l + \kappa^{d_l\star}_m - \kappa^{s_l\star}_m = 0,\ \forall\, m,\ \forall\, l \in \mathcal{L} \tag{C.3d}$$

$$0 \le \theta^\star_l \perp \bar{f}_l(\mathbf{v}^\star) - \sum_{m=1}^{M} f^\star_{l,m} \ge 0,\ \forall\, l \in \mathcal{L}^{wl}. \tag{C.3e}$$

For problem (C.2), we have

$$-2\hat{\epsilon}_{s_l}\hat{v}_l - \sum_{n\in\bar{I}(l)} \hat{\theta}_n \hat{w}_n \nabla_{v_l} e_n(\hat{u}_n, \hat{\mathbf{v}}) = 0, \tag{C.4a}$$

$$\hat{\theta}_l(\nabla_{u_l} e_l(\hat{u}_l, \hat{\mathbf{v}})) = 0,\ \hat{\theta}_l \left(\frac{1}{w_l} - e_l(\hat{u}_l, \hat{\mathbf{v}})\right) = 0,\ \forall\, l \in \mathcal{L}^{wl}, \tag{C.4b}$$

$$-\hat{\delta} + \sum_{m=1}^{M} \hat{\delta}_m = 1, \tag{C.4c}$$

$$\hat{\delta}_m + \hat{\kappa}^{s_m}_m - \hat{\kappa}^{d_m}_m = 0,\ \forall\, m, \tag{C.4d}$$

$$\hat{\delta}_{l,m} - \hat{\theta}_l + \hat{\kappa}^{d_l}_m - \hat{\kappa}^{s_l}_m = 0,\ \forall\, m,\ \forall\, l \in \mathcal{L} \tag{C.4e}$$

$$0 \le \hat{\theta}_l \perp \left(1 + \log(\hat{w}_l) - \hat{w}_l e_l(\hat{u}_l, \hat{\mathbf{v}}) - \sum_{m=1}^{M} \hat{f}_{l,m}\right) \ge 0,\ \forall\, l \in \mathcal{L}^{wl}. \tag{C.4f}$$

Obviously, by comparing (C.3b)∼(C.3d) and (C.4c)∼(C.4e), we can conclude that $\hat{\boldsymbol{y}}$ can satisfy (C.4c)∼(C.4e). For (C.4b), by the optimality of (3.8) and (3.9), they are also true for $\hat{\boldsymbol{y}}$. Moreover, it follows from Lemma 2 that

$$\begin{aligned}\bar{f}_l(\mathbf{v}^\star) &= 1 + \log(w_l(\mathbf{v}^\star)) - w_l(\mathbf{v}^\star)e_l(u_l(\mathbf{v}^\star), \mathbf{v}^\star) \\ &= 1 + \log(\hat{w}_l) - \hat{w}_l e_l(\hat{u}_l, \mathbf{v}^\star),\end{aligned} \tag{C.5}$$

with this fact and by (C.3e), (C.4f) is satisfied with $\hat{\boldsymbol{y}}$.

For the last KKT condition of problem (C.2) that the relationship has not been established, i.e., (C.4a), let us first split the Lagrange multiplier $\boldsymbol{\theta}^\star$ of problem (C.1) into two subsets $\mathcal{A} \triangleq \{l \mid \theta_l^\star > 0,\ \forall\ l \in \mathcal{L}\}$ and $\bar{\mathcal{A}} \triangleq \{l \mid \theta_l^\star = 0,\ \forall\ l \in \mathcal{L}\}$. In the following, we will only consider the set of wireless capacity constraints belonging to $\mathcal{A}$, i.e., the active wireless capacity constraints. Note that, for these constraints, variables $\hat{u}_l$ and $\hat{w}_l$ can be uniquely determined by (C.4b), but those belong to $\bar{\mathcal{A}}$ cannot be uniquely determined. By exploiting these facts, KKT conditions (C.3a) and (C.4a) for $\forall\ l \in \mathcal{L}^{wl}$ can be related as follows

$$\begin{aligned}\text{(C.4a)} =& -2\hat{\epsilon}_{s_l}\hat{v}_l - \sum_{n\in\bar{I}(l)} \hat{\theta}_n\hat{w}_n\nabla_{v_l}e_n(\hat{u}_n, \hat{\mathbf{v}}) \\ =& -2\hat{\epsilon}_{s_l}\hat{v}_l - \sum_{n\in\bar{I}(l),\ n\in\mathcal{A}} \hat{\theta}_n\hat{w}_n\nabla_{v_l}e_n(\hat{u}_n, \hat{\mathbf{v}}) \\ =& -2\epsilon_{s_l}^\star v_l^\star + \sum_{n\in\bar{I}(l),\ n\in\mathcal{A}} \theta_n^\star\nabla_{v_l}\bar{f}_n(\mathbf{v}^\star) \qquad \text{(by (C.5))} \\ =& -2\epsilon_{s_l}^\star v_l^\star + \sum_{n\in\bar{I}(l)} \theta_n^\star\nabla_{v_l}\bar{f}_n(\mathbf{v}^\star) = \text{(C.3a)} = 0.\end{aligned}$$

Therefore, (C.4a) is also satisfied by $\hat{\boldsymbol{y}}$, so it is a stationary solution of problem (C.2). The reverse statement of Step 1 can be argued similarly.

Step 2: Every global optimal solution of problem (C.1) **corresponds to a global optimal solution of problem** (C.2) **with the same objective value.**

In the following, we first show that if $y^\star \triangleq \{\mathbf{r}^\star, \mathbf{v}^\star, \tilde{\mathbf{u}}, \tilde{\mathbf{w}}; \boldsymbol{\delta}^\star, \boldsymbol{\theta}^\star, \boldsymbol{\kappa}^\star, \boldsymbol{\epsilon}^\star\}$ is an arbitrarily KKT solution of problem (C.2), $\hat{y} \triangleq \{\mathbf{r}^\star, \mathbf{v}^\star, \mathbf{u}(\mathbf{v}^\star), \mathbf{w}(\mathbf{v}^\star); \boldsymbol{\delta}^\star, \boldsymbol{\theta}^\star, \boldsymbol{\kappa}^\star, \boldsymbol{\epsilon}^\star\}$ is also a KKT solution of problem (C.2). Furthermore, both of them achieve the same objective value. Note that, the difference between $y^\star$ and $\hat{y}$, i.e., $\mathbf{u}$ and $\boldsymbol{w}$, appear in KKT

conditions (C.4a), (C.4b), and (C.4f). In the following, we will check each of them for $\hat{y}$.

For (C.4b): Observe that for $l \in \mathcal{A}$, the condition implies

$$\nabla_{u_l} e_l(\tilde{u}_l, \mathbf{v}^\star) = 0,\ \frac{1}{w_l} - e_l(\tilde{u}_l, \mathbf{v}^\star) = 0,$$

and it reduces to the fact that

$$\tilde{u}_l = u_l(\mathbf{v}^\star),\ \text{and}\ \tilde{w}_l = w_l(\mathbf{v}^\star). \tag{C.6}$$

For $l \in \bar{\mathcal{A}}$, any choice of u_l and w_l would always satisfy condition (C.4b).

For (C.4a): $\hat{y}$ satisfies (C.4a) can be derived as follows:

$$\begin{aligned}
0 =& -2\epsilon^\star_{s_l} v^\star_l - \sum_{n\in\bar{I}(l)} \theta^\star_n w^\star_n \nabla_{v_l} e_n(\tilde{u}_n, \mathbf{v}^\star) \\
=& -2\epsilon^\star_{s_l} v^\star_l - \sum_{n\in\bar{I}(l),\ n\in\mathcal{A}} \theta^\star_n w^\star_n \nabla_{v_l} e_n(\tilde{u}_n, \mathbf{v}^\star) \\
\underbrace{=}_{(a)}& -2\epsilon^\star_{s_l} v^\star_l - \sum_{n\in\bar{I}(l),\ n\in\mathcal{A}} \theta^\star_n w^\star_n \nabla_{v_l} e_n(u_n(\mathbf{v}^\star), \mathbf{v}^\star) \\
=& -2\epsilon^\star_{s_l} v^\star_l - \sum_{n\in\bar{I}(l)} \theta^\star_n w^\star_n \nabla_{v_l} e_n(u_n(\mathbf{v}^\star), \mathbf{v}^\star),
\end{aligned}$$

where (a) is due to the fact that when $l \in \mathcal{A}$, (C.6) holds. Hence, KKT condition (C.4a) is also satisfied by $\hat{y}$.

For (C.4f): Due to the update rule choice for $\mathbf{u}$ and $\boldsymbol{w}$, i.e., (3.8) and (3.9), we have the following conclusion,

$$\begin{aligned}
&\left(1 + \log(w_l(\mathbf{v}^\star)) - w_l(\mathbf{v}^\star) e_l(u_l(\mathbf{v}^\star), \mathbf{v}^\star) - \sum_{m=1}^{M} f^\star_{l,m}\right) \\
\geq &\left(1 + \log(\tilde{w}_l) - \tilde{w}_l e_l(\tilde{u}_l, \mathbf{v}^\star) - \sum_{m=1}^{M} f^\star_{l,m}\right) \geq 0.
\end{aligned}$$

This result implies that feasibility part of (C.4f) is satisfied. In order to show that the complementarity part is also satisfied, it is sufficient to show that for all $l \in \mathcal{A}$, the above inequality achieves strict equality since (C.6) holds at this scenario.

In sum, we can conclude that $\hat{y}$ is also a KKT solution of problem (C.2). Recall that the network is connected and the link capacities are all positive. Hence, the optimal

value $r^\star_{\min}$ must be strictly greater than 0. It follows that the Lagrangian dual variable for constraint $r^\star_{\min} \geq 0$, i.e., $\delta^\star$, is always 0 by the complementarity condition. The argument follows that we must have $|\mathcal{A}| > 0$, and at least one of the constraints $r^\star_m \geq r^\star_{\min}$, $m = 1 \sim M$, is active. Since $r^\star_{\min}$ is the objective value for both KKT solutions $y^\star$ and $\hat{y}$, they both achieves the same objective value.

Now we are ready to argue that if $y^\star = \{\mathbf{r}^\star, \mathbf{v}^\star, \mathbf{u}(\mathbf{v}^\star), \boldsymbol{w}(\mathbf{v}^\star)\}$ is the optimal solution of (C.2), the objective value should be the same as the optimal value of problem (C.1). To prove this, we will resort to contradiction. In particular, the $r^\star_{\min}$ achieved by $\{\mathbf{r}^\star, \mathbf{v}^\star\}$ of $y^\star$ for problem (C.1) is the same as $r^\star_{\min}$ achieved by $\boldsymbol{y}^\star$ for problem (C.2) by applying Lemma 2. Assume $r^\star_{\min}$ is not the optimal objective value of problem (C.1), and the optimal solution of problem (C.1) is $\hat{\boldsymbol{x}} = \{\hat{\mathbf{r}}, \hat{\mathbf{v}}\}$ with objective value $\hat{r}_{\min} > r^\star_{\min}$. Since $\hat{r}_{\min}$ can also achieved for problem (C.2) by $\hat{y} = \{\hat{\mathbf{r}}, \hat{\mathbf{v}}, \mathbf{u}(\hat{\mathbf{v}}), \boldsymbol{w}(\hat{\mathbf{v}})\}$ when we apply Lemma 2 again. Hence, the optimality of $\boldsymbol{y}^\star$ is violated by the existence of $\hat{\boldsymbol{y}}$. This contradiction concludes the optimal value for problem (C.1) is also $r^\star_{\min}$. The reverse direction can be argued similarly.

Step 3: The N-MaxMin Algorithm can converge to the KKT solutions of problem (C.1)**.**

Since the objective value generated by the proposed N-MaxMin Algorithm for problem (C.2) is monotonically increasing, and the objective value of problem (C.2) is finite. Hence, the generated sequence of objective value converges. Due to the compactness of the feasible set for problem (C.2), the iterates $\{\mathbf{r}^{(t)}, \mathbf{v}^{(t)}\}$ must have a cluster point $\{\bar{\mathbf{r}}, \bar{\mathbf{v}}\}$. Let $\{\mathbf{r}^{n_t}, \mathbf{v}^{n_t}\}_{t=1}^{\infty}$ be the subsequence converging to $\{\bar{\mathbf{r}}, \bar{\mathbf{v}}\}$. Since the maps $\mathbf{u}(\mathbf{v})$ and $\boldsymbol{w}(\mathbf{v})$ are continuous, we must have

$$\lim_{t\to\infty} (\mathbf{r}^{n_t}, \mathbf{v}^{n_t}, \mathbf{u}^{n_t}, \boldsymbol{w}^{n_t}) = (\bar{\mathbf{r}}, \bar{\mathbf{v}}, \bar{\mathbf{u}}, \bar{\boldsymbol{w}}) \triangleq (\bar{\mathbf{r}}, \bar{\mathbf{v}}, \mathbf{u}(\mathbf{v}), \boldsymbol{w}(\mathbf{v})).$$

First we will show that in the limit we have: $\{\bar{\mathbf{r}}, \bar{\mathbf{v}}\} \in \Phi(\bar{\mathbf{u}}, \bar{\boldsymbol{w}})$ where $\Phi(\mathbf{u}, \boldsymbol{w})$ is the mapping from given $\mathbf{u}$ and $\boldsymbol{w}$ to the optimal solution for problem (C.2). Due to the optimality of $\{\mathbf{r}^{n_t}, \mathbf{v}^{n_t}\}$ and the monotonic increase of the objective function, we have $\{\bar{\mathbf{r}}, \bar{\mathbf{v}}\} \in \Phi(\bar{\mathbf{u}}, \bar{\boldsymbol{w}})$.

The next step is to establish that $\{\bar{\mathbf{r}}, \bar{\mathbf{v}}, \bar{\mathbf{u}}, \bar{\boldsymbol{w}}\} = \{\bar{\mathbf{r}}, \bar{\mathbf{v}}, \mathbf{u}(\bar{\mathbf{v}}), \boldsymbol{w}(\bar{\mathbf{v}})\}$ is a KKT solution of (C.2). This is true due to the two facts i) $\{\bar{\mathbf{r}}, \bar{\mathbf{v}}\} \in \Phi(\bar{\mathbf{u}}, \bar{\boldsymbol{w}})$; and ii) $\bar{\mathbf{u}} = \mathbf{u}(\bar{\mathbf{v}})$ and $\bar{\boldsymbol{w}} = \boldsymbol{w}(\bar{\boldsymbol{w}})$. Using these two facts, the KKT conditions for problem (C.2), i.e., (C.4),

are satisfied. Applying the result of Step 1, we conclude that $\{\bar{\mathbf{r}}, \bar{\mathbf{v}}\}$ must be a KKT point of the original problem (C.1).

So far we have proved that any cluster point of the iterates is a KKT point of problem (C.1). Since the feasible set is compact, we have N-MaxMin Algorithm can converge to the KKT solutions. □

Appendix D

Proof of Theorem 5

We first derive some useful properties of the proposed PAsyn-BSUM-M algorithm. For notational simplicity, we will use the shorthand $\mathbf{z} = [\boldsymbol{x}^T, \boldsymbol{\lambda}^T]^T$. In the first step, we will characterize the successive difference of the augmented Lagrangian over iterations, i.e., $L(\mathbf{z}^t) - L(\mathbf{z}^{t+1})$.

Lemma 3 *For the PAsyn-BSUM-M algorithm, we have*

$$\begin{aligned} & L(\mathbf{z}^t) - L(\mathbf{z}^{t+1}) \\ \geq & \sum_{i=0}^{K} I_{p,i}^t \frac{2\gamma + \rho\lambda_{\min}\left(\sum_{j=1}^{J} \tilde{\mathbf{A}}_{ij}^T \tilde{\mathbf{A}}_{ij}\right)}{2} \|\boldsymbol{x}_i^t - \boldsymbol{x}_i^{t+1}\|^2 + \sum_{j=1}^{J} I_{d,j}^t \langle \tilde{\mathbf{A}}_j \boldsymbol{x}^{t+1} - \tilde{\mathbf{b}}_j, \boldsymbol{\lambda}_j^t - \boldsymbol{\lambda}_j^{t+1} \rangle. \end{aligned} \tag{D.1}$$

Proof: We first decompose $L(\mathbf{z}^t) - L(\mathbf{z}^{t+1})$ as

$$\begin{aligned} & L(\mathbf{z}^t) - L(\mathbf{z}^{t+1}) \\ = & \sum_{i=0}^{K} \left[L(\boldsymbol{x}_0^{t+1}, \cdots, \boldsymbol{x}_{i-1}^{t+1}, \boldsymbol{x}_i^t, \cdots, \boldsymbol{x}_K^t; \boldsymbol{\lambda}^t) - L(\boldsymbol{x}_0^{t+1}, \cdots, \boldsymbol{x}_i^{t+1}, \boldsymbol{x}_{i+1}^t, \cdots, \boldsymbol{x}_K^t; \boldsymbol{\lambda}^t) \right] \\ + & \sum_{j=1}^{J} \left[L(\boldsymbol{x}^{t+1}; \boldsymbol{\lambda}_1^{t+1}, \cdots, \boldsymbol{\lambda}_{j-1}^{t+1}, \boldsymbol{\lambda}_j^t, \cdots, \boldsymbol{\lambda}_J^t) - L(\boldsymbol{x}^{t+1}; \boldsymbol{\lambda}_1^{t+1}, \cdots, \boldsymbol{\lambda}_j^{t+1}, \boldsymbol{\lambda}_{j+1}^t, \cdots, \boldsymbol{\lambda}_J^t) \right]. \end{aligned}$$

In the sequel, we lower bound each term pair individually.

(Case 1) Only $\boldsymbol{x}_i$ is updated at the tth iteration, i.e., $I_{p,i}^t = 1$:

$$\begin{aligned}
&L(\mathbf{z}^t) - L(\mathbf{z}^{t+1}) \\
=&L(\boldsymbol{x}_0^{t+1}, \cdots, \boldsymbol{x}_{i-1}^{t+1}, \boldsymbol{x}_i^t, \cdots, \boldsymbol{x}_K^t; \boldsymbol{\lambda}^t) - L(\boldsymbol{x}_0^{t+1}, \cdots, \boldsymbol{x}_i^{t+1}, \boldsymbol{x}_{i+1}^t, \cdots, \boldsymbol{x}_K^t; \boldsymbol{\lambda}^t) \\
\geq&\langle \nabla_{\boldsymbol{x}_i} L(\boldsymbol{x}_0^{t+1}, \cdots, \boldsymbol{x}_i^{t+1}, \boldsymbol{x}_{i+1}^t, \cdots, \boldsymbol{x}_K^t; \boldsymbol{\lambda}^t), \boldsymbol{x}_i^t - \boldsymbol{x}_i^{t+1} \rangle \\
&+ \frac{\rho \lambda_{\min}\left(\sum_{j=1}^J \tilde{\mathbf{A}}_{ij}^T \tilde{\mathbf{A}}_{ij}\right)}{2} \|\boldsymbol{x}_i^t - \boldsymbol{x}_i^{t+1}\|^2. \qquad \text{(D.2)}
\end{aligned}$$

The inequality is due to the fact that $\nabla_{\boldsymbol{x}_0}^2 L(\mathbf{z}) \succeq \rho \sum_{j=1}^J \tilde{\mathbf{A}}_{ij}^T \tilde{\mathbf{A}}_{ij}$. Since $\boldsymbol{x}_i^{t+1}$ is the optimal solution of the following convex optimization,

$$\boldsymbol{x}_i^{t+1} = \arg\min_{\boldsymbol{x}_i \in \mathcal{X}_i} L(\boldsymbol{x}_0^{t+1}, \cdots, \boldsymbol{x}_{i-1}^{t+1}, \boldsymbol{x}_i, \boldsymbol{x}_{i+1}^t, \cdots, \boldsymbol{x}_K^t; \boldsymbol{\lambda}^t) + \frac{\gamma}{2} \|\boldsymbol{x}_i - \boldsymbol{x}_i^t\|^2. \qquad \text{(D.3)}$$

Hence, by the first-order optimality condition, we have

$$\begin{aligned}
&\langle \nabla_{\boldsymbol{x}_i} L(\boldsymbol{x}_0^{t+1}, \cdots, \boldsymbol{x}_i^{t+1}, \boldsymbol{x}_{i+1}^t, \cdots, \boldsymbol{x}_K^t; \boldsymbol{\lambda}^t) + \gamma(\boldsymbol{x}_i^{t+1} - \boldsymbol{x}_i^t), \boldsymbol{x}_i^t - \boldsymbol{x}_i^{t+1} \rangle \geq 0 \\
\Rightarrow&\langle \nabla_{\boldsymbol{x}_i} L(\boldsymbol{x}_0^{t+1}, \cdots, \boldsymbol{x}_i^{t+1}, \boldsymbol{x}_{i+1}^t, \cdots, \boldsymbol{x}_K^t; \boldsymbol{\lambda}^t), \boldsymbol{x}_i^t - \boldsymbol{x}_i^{t+1} \rangle \geq \gamma \|\boldsymbol{x}_i^{t+1} - \boldsymbol{x}_i^t\|^2. \qquad \text{(D.4)}
\end{aligned}$$

Substitute (D.4) into (D.2), we conclude that

$$L(\mathbf{z}^t) - L(\mathbf{z}^{t+1}) \geq \frac{2\gamma + \rho \lambda_{\min}\left(\sum_{j=1}^J \tilde{\mathbf{A}}_{ij}^T \tilde{\mathbf{A}}_{ij}\right)}{2} \|\boldsymbol{x}_i^t - \boldsymbol{x}_i^{t+1}\|^2. \qquad \text{(D.5)}$$

(Case 2) Only $\boldsymbol{\lambda}_j$ is updated at the tth iteration, i.e., $I_{d,j}^t = 1$: For this case, we can straightforwardly obtain the following equality relationship,

$$\begin{aligned}
&L(\mathbf{z}^t) - L(\mathbf{z}^{t+1}) \\
=&L(\boldsymbol{x}^{t+1}; \boldsymbol{\lambda}_1^{t+1}, \cdots, \boldsymbol{\lambda}_{j-1}^{t+1}, \boldsymbol{\lambda}_j^t, \cdots, \boldsymbol{\lambda}_J^t) - L(\boldsymbol{x}^{t+1}; \boldsymbol{\lambda}_1^{t+1}, \cdots, \boldsymbol{\lambda}_j^{t+1}, \boldsymbol{\lambda}_{j+1}^t, \cdots, \boldsymbol{\lambda}_J^t) \\
=&\langle \tilde{\mathbf{A}}_j \boldsymbol{x}^{t+1}, \boldsymbol{\lambda}_j^t - \boldsymbol{\lambda}_j^{t+1} \rangle. \qquad \text{(D.6)}
\end{aligned}$$

By combining both case 1 and case 2, i.e., (D.5) and (D.6), we obtain the desired result (D.1).

□

We introduce the proximal gradient, which will be used as a measure of optimality.

Definition 1 *(Proximal gradient) Let $\mathcal{X} \subseteq \mathbb{R}^n$ be a nonempty closed convex set. The proximal gradient of convex and smooth function f is defined as*

$$\tilde{\nabla} f(\boldsymbol{x}) \triangleq \boldsymbol{x} - \mathrm{Proj}_{\mathcal{X}}(\boldsymbol{x} - \nabla f(x))$$

With this definition of the proximal gradient, in the next step, the norm of it at any given iterate $t+1$ can be upper bounded by the following lemma.

Lemma 4 *The gradient of the augmented Lagrangian function* $\|\tilde{\nabla}L(\boldsymbol{x}^{t+1};\boldsymbol{\lambda}^{t+1})\|$ *can be upper bounded as:*

$$\|\tilde{\nabla}L(\boldsymbol{x}^{t+1};\boldsymbol{\lambda}^{t+1})\|^2 \leq \sigma \left[\|\boldsymbol{x}^{t+1}-\boldsymbol{x}^{t}\|^2 + \|\boldsymbol{x}^{t}-\boldsymbol{x}^{t-1}\|^2 + \sum_{j=1}^{J} \|\boldsymbol{\lambda}_j^{t+1}-\boldsymbol{\lambda}_j^{t}\|^2 \right], \quad \text{(D.7)}$$

where σ *are some fixed positive constant.*

Proof: By triangular inequality, the gradient $\|\tilde{\nabla}L(\boldsymbol{x}^{t+1};\boldsymbol{\lambda}^{t+1})\|^2$ can be upper bounded with

$$\|\tilde{\nabla}L(\boldsymbol{x}^{t+1};\boldsymbol{\lambda}^{t+1})\|^2 \leq \left(\sum_{i=0}^{K} \|\tilde{\nabla}_{\boldsymbol{x}_i}L(\boldsymbol{x}^{t+1};\boldsymbol{\lambda}^{t+1})\| \right)^2. \quad \text{(D.8)}$$

In the following, we will bound each term individually. The upper bound of

$$\|\tilde{\nabla}_{\boldsymbol{x}_i}L(\boldsymbol{x}^{t+1};\boldsymbol{\lambda}^{t+1})\|,\ i=0\sim K,$$

can be derived as follows:

$$\begin{aligned}
&\|\tilde{\nabla}_{\boldsymbol{x}_i}L(\boldsymbol{x}^{t+1};\boldsymbol{\lambda}^{t+1})\| \leq \|\boldsymbol{x}_i^{t+1}-\boldsymbol{x}_i^{t}\| + \|\boldsymbol{x}_i^{t} - \mathrm{Proj}_{\mathcal{X}_i}\left(\boldsymbol{x}_i^{t+1} - \nabla_{\boldsymbol{x}_i}L(\boldsymbol{x}^{t+1};\boldsymbol{\lambda}^{t+1})\right)\| \\
=&\|\boldsymbol{x}_i^{t+1}-\boldsymbol{x}_i^{t}\| \\
&+ \|\mathrm{Proj}_{\mathcal{X}_i}(\boldsymbol{x}_i^{t} - \nabla_{\boldsymbol{x}_i}L(\boldsymbol{x}^{t};\boldsymbol{\lambda}^{t}) - \gamma(\boldsymbol{x}_i^{t}-\boldsymbol{x}_i^{t-1})) - \mathrm{Proj}_{\mathcal{X}_i}(\boldsymbol{x}_i^{t+1} - \nabla_{\boldsymbol{x}_i}L(\boldsymbol{x}^{t+1};\boldsymbol{\lambda}^{t+1})\| \\
\leq&\|\boldsymbol{x}_i^{t+1}-\boldsymbol{x}_i^{t}\| + \left\| \left\{ \boldsymbol{x}_i^{t} - \left[\nabla_{\boldsymbol{x}_i} f_i(\boldsymbol{x}_i^{t}) + \sum_{j=1}^{J} \tilde{\mathbf{A}}_{ij}^T \left[\boldsymbol{\lambda}_j^{t} + \rho(\tilde{\mathbf{A}}_j\boldsymbol{x}^{t} - \tilde{\mathbf{b}}_j) \right] + \gamma(\boldsymbol{x}_i^{t}-\boldsymbol{x}_i^{t-1}) \right] \right\} \right. \\
&\qquad \left. - \left\{ \boldsymbol{x}_i^{t+1} - \left[\nabla f_i(\boldsymbol{x}_i^{t+1}) + \sum_{j=1}^{J} \mathbf{A}_{ij}^T \left[\boldsymbol{\lambda}_j^{t+1} + \rho(\tilde{\mathbf{A}}_j\boldsymbol{x}^{t+1} - \tilde{\mathbf{b}}_j) \right] \right] \right\} \right\| \\
\leq&(2+L)\|\boldsymbol{x}_i^{t+1}-\boldsymbol{x}_i^{t}\| + \gamma\|\boldsymbol{x}_i^{t}-\boldsymbol{x}_i^{t-1}\| \\
&+ \|\boldsymbol{x}^{t+1}-\boldsymbol{x}^{t}\| \sum_{j=1}^{J} \rho\|\tilde{\mathbf{A}}_{ij}^T\tilde{\mathbf{A}}_j\| + \sum_{j=1}^{J} \|\tilde{\mathbf{A}}_{ij}^T\|\|\boldsymbol{\lambda}_j^{t+1}-\boldsymbol{\lambda}_j^{t}\| \qquad \text{(D.9)}
\end{aligned}$$

By summing these individual proximal gradient upper bounds for each $i=0\sim K$, and use the fact that $(\sum_{i=1}^{K}\|\mathbf{a}_i\|)^2 \leq K\sum_{i=1}^{K}\|\mathbf{a}_i\|^2$ for arbitrary $\{\mathbf{a}_i\}_i$. There is a positive constant σ such that the desired result (D.7) holds. □

To analyze the convergence of the algorithms by measuring the algorithm progress, we need to make use of a certain "potential function". Here, the summation of the *dual optimality gap* and *primal optimality gap* is adopted, each of which is defined below:

- Dual optimality gap: $\triangle_d^t = d^\star - d(\boldsymbol{\lambda}^t) \geq 0$,
- Primal optimality gap: $\triangle_p^t = L(\mathbf{z}^t) - d(\boldsymbol{\lambda}^t) \geq 0$,

where $d^\star$ is the optimal value for the dual problem of (4.1), i.e., $d^\star \triangleq \max_{\boldsymbol{\lambda}} \min_{\boldsymbol{x}\in\mathcal{X}} L(\boldsymbol{x};\boldsymbol{\lambda})$, and $d(\boldsymbol{\lambda}^t) \triangleq \min_{\boldsymbol{x}\in\mathcal{X}} L(\boldsymbol{x};\boldsymbol{\lambda}^t)$ with optimal $\boldsymbol{x}$ denoted as $\bar{\boldsymbol{x}}^t$

In the next lemma, the upper bound of the potential function over iteration is derived.

Lemma 5 *For the PAsyn-BSUM-M algorithm, there holds*

$$\triangle_p^{t+1} + \triangle_d^{t+1} - (\triangle_p^t + \triangle_d^t) \leq L(\mathbf{z}^{t+1}) - L(\mathbf{z}^t) + 2\sum_{j=1}^{J} I_{d,j}^t \langle \tilde{\mathbf{A}}_j \bar{\boldsymbol{x}}^{t+1} - \tilde{\mathbf{b}}_j, \boldsymbol{\lambda}_j^t - \boldsymbol{\lambda}_j^{t+1}\rangle. \tag{D.10}$$

Proof: The upper bound can be straightforwardly derived by the following procedures:

$$\begin{aligned}
&\triangle_p^{t+1} + \triangle_d^{t+1} - (\triangle_p^t + \triangle_d^t) \\
=&L(\mathbf{z}^{t+1}) - L(\mathbf{z}^t) + 2(d(\boldsymbol{\lambda}^t) - d(\boldsymbol{\lambda}^{t+1})) \\
=&L(\mathbf{z}^{t+1}) - L(\mathbf{z}^t) + 2(L(\bar{\boldsymbol{x}}^t;\lambda^t) - L(\bar{\boldsymbol{x}}^{t+1};\boldsymbol{\lambda}^{t+1})) \\
=&L(\mathbf{z}^{t+1}) - L(\mathbf{z}^t) + 2[(L(\bar{\boldsymbol{x}}^{t+1};\boldsymbol{\lambda}^t) - L(\bar{\boldsymbol{x}}^{t+1};\boldsymbol{\lambda}^{t+1})) + (L(\bar{\boldsymbol{x}}^t;\boldsymbol{\lambda}^t) - L(\bar{\boldsymbol{x}}^{t+1};\boldsymbol{\lambda}^t))] \\
\leq&L(\mathbf{z}^{t+1}) - L(\mathbf{z}^t) + 2\sum_{j=1}^{J} I_{d,j}^t \langle \tilde{\mathbf{A}}_j \bar{\boldsymbol{x}}^{t+1} - \tilde{\mathbf{b}}_j, \boldsymbol{\lambda}_j^t - \boldsymbol{\lambda}_j^{t+1}\rangle,
\end{aligned} \tag{D.11}$$

where the inequality is due to the fact that $L(\bar{\boldsymbol{x}}^t;\boldsymbol{\lambda}^t) \leq L(\bar{\boldsymbol{x}}^{t+1};\boldsymbol{\lambda}^t)$ by the definition of $\bar{\boldsymbol{x}}^t$. □

Given the previous necessary properties on the upper bound of the considered potential function, we are ready for performing the final convergence analysis of the proposed

PAsyn-BSUM-M algorithm. By Lemma 3 and 5, we can conclude that

$$\begin{aligned}
&\left[\triangle_p^{t+1}+\triangle_d^{t+1}-(\triangle_p^t+\triangle_d^t)\right]\\
\leq&-\sum_{i=0}^{K} I_{p,i}^t \frac{2\gamma+\rho\lambda_{\min}\left(\sum_{j=1}^{J}\tilde{\mathbf{A}}_{ij}^T\tilde{\mathbf{A}}_{ij}\right)}{2}\|\boldsymbol{x}_i^t-\boldsymbol{x}_i^{t+1}\|^2\\
&+\sum_{j=1}^{J} I_{d,j}^t \underbrace{\left[\langle\tilde{\mathbf{A}}_j\boldsymbol{x}^{t+1}-\tilde{\mathbf{b}}_j,\boldsymbol{\lambda}_j^{t+1}-\boldsymbol{\lambda}_j^t\rangle-2\langle\tilde{\mathbf{A}}_j\bar{\boldsymbol{x}}^{t+1}-\tilde{\mathbf{b}}_j,\boldsymbol{\lambda}_j^{t+1}-\boldsymbol{\lambda}_j^t\rangle\right]}_{a_j^{t+1}}. \qquad \text{(D.12)}
\end{aligned}$$

Let us first upper bound term a_j^{t+1},

$$\begin{aligned}
a_j^{t+1}&=\alpha^t\Big[\langle\tilde{\mathbf{A}}_j\boldsymbol{x}^{t+1}-\tilde{\mathbf{b}}_j,\tilde{\mathbf{A}}_j\boldsymbol{x}^{t+1}-\tilde{\mathbf{b}}_j\rangle-2\langle\tilde{\mathbf{A}}_j\bar{\boldsymbol{x}}^{t+1}-\tilde{\mathbf{b}}_j,\tilde{\mathbf{A}}_j\boldsymbol{x}^{t+1}-\tilde{\mathbf{b}}_j\rangle\\
&\qquad+\langle\tilde{\mathbf{A}}_j\boldsymbol{x}^{t+1}-\tilde{\mathbf{b}}_j,\tilde{\mathbf{A}}_j\boldsymbol{x}^{t+1}-\tilde{\mathbf{b}}_j\rangle-\langle\tilde{\mathbf{A}}_j\boldsymbol{x}^{t+1}-\tilde{\mathbf{b}}_j,\tilde{\mathbf{A}}_j\boldsymbol{x}^{t+1}-\tilde{\mathbf{b}}_j\rangle\Big]\\
&=\alpha^t\left[\|\tilde{\mathbf{A}}_j(\boldsymbol{x}^{t+1}-\bar{\boldsymbol{x}}^{t+1})\|^2-\|\tilde{\mathbf{A}}_j\boldsymbol{x}^{t+1}-\tilde{\mathbf{b}}_j\|^2\right], \qquad \text{(D.13)}
\end{aligned}$$

where the first equality is due to the update rule that when $I_{d,j}^t=1$, $\boldsymbol{x}^t=\boldsymbol{x}^{t+1}$ and $\boldsymbol{\lambda}_j^{t+1}-\boldsymbol{\lambda}_j^t=\alpha^t(\tilde{\mathbf{A}}_j\boldsymbol{x}^{t+1}-\tilde{\mathbf{b}}_j)$. In the following, the local error bound property will be adopted to upper bound $\|\tilde{\mathbf{A}}_j(\boldsymbol{x}^{t+1}-\bar{\boldsymbol{x}}^{t+1})\|^2$. Specifically, since the assumptions (A3) and (A4) satisfy the conditions of [108, Lemma 2.2], the following result can hence be exploited:

$$\|\boldsymbol{x}^{t+1}-\bar{\boldsymbol{x}}^{t+1}\|\leq\tilde{\tau}\|\tilde{\nabla}L(\boldsymbol{x}^{t+1};\boldsymbol{\lambda}^{t+1})\|, \qquad \text{(D.14)}$$

where $\tilde{\tau}$ is some positive constant. By applying this local error bound property,

$$\begin{aligned}
&\|\tilde{\mathbf{A}}_j(\boldsymbol{x}^{t+1}-\bar{\boldsymbol{x}}^{t+1})\|^2\leq\|\tilde{\mathbf{A}}_j\|^2\|\boldsymbol{x}^{t+1}-\bar{\boldsymbol{x}}^{t+1}\|^2\leq\|\tilde{\mathbf{A}}_j\|^2\tilde{\tau}^2\|\tilde{\nabla}L(\boldsymbol{x}^{t+1};\boldsymbol{\lambda}^{t+1})\|^2\\
&\leq\|\tilde{\mathbf{A}}_j\|^2\tilde{\tau}^2\sigma\left[\|\boldsymbol{x}^{t+1}-\boldsymbol{x}^t\|^2+\|\boldsymbol{x}^t-\boldsymbol{x}^{t-1}\|^2+\sum_{j=1}^{J}\|\boldsymbol{\lambda}_j^{t+1}-\boldsymbol{\lambda}_j^t\|^2\right]\\
&\leq\|\tilde{\mathbf{A}}_j\|^2\left[\underbrace{\tilde{\tau}^2\sigma\left(\|\boldsymbol{x}^{t+1}-\boldsymbol{x}^t\|^2+\|\boldsymbol{x}^t-\boldsymbol{x}^{t-1}\|^2\right)}_{z_0^{t+1}}+\underbrace{\tilde{\tau}^2\sigma\alpha^t}_{z_1^{t+1}}\sum_{j=1}^{J}\|\tilde{\mathbf{A}}_j\boldsymbol{x}^{t+1}-\tilde{\mathbf{b}}_j\|^2\right], \qquad \text{(D.15)}
\end{aligned}$$

where the third inequality is due to (D.7).

Moreover, the upper bound for $\sum_{j=1}^{J}\|\tilde{\mathbf{A}}_j\boldsymbol{x}^{t+1}-\tilde{\mathbf{b}}_j\|^2$ can be derived by (D.15) and adopting the property of local error bound lemma [108, Lemma 2.2] again. Specifically,

$$\begin{aligned}
&\sum_{j=1}^{J}\|\tilde{\mathbf{A}}_j\boldsymbol{x}^{t+1}-\tilde{\mathbf{b}}_j\|^2=\sum_{j=1}^{J}\|\tilde{\mathbf{A}}_j(\boldsymbol{x}^{t+1}-\bar{\boldsymbol{x}}^{t+1})+\tilde{\mathbf{A}}_j\bar{\boldsymbol{x}}^{t+1}-\tilde{\mathbf{b}}_j\|^2\\
\leq&2\sum_{j=1}^{J}\left[\|\tilde{\mathbf{A}}_j(\boldsymbol{x}^{t+1}-\bar{\boldsymbol{x}}^{t+1})\|^2+\|\tilde{\mathbf{A}}_j\bar{\boldsymbol{x}}^{t+1}-\tilde{\mathbf{b}}_j\|^2\right]\\
\leq&2\sum_{j=1}^{J}\left(z_0^{t+1}\|\tilde{\mathbf{A}}_j\|^2+\|\tilde{\mathbf{A}}_j\bar{\boldsymbol{x}}^{t+1}-\tilde{\mathbf{b}}_j\|^2\right)+\left(2z_1^{t+1}\sum_{j=1}^{J}\|\tilde{\mathbf{A}}_j\|^2\right)\left(\sum_{j=1}^{J}\|\tilde{\mathbf{A}}_j\boldsymbol{x}^{t+1}-\tilde{\mathbf{b}}_j\|^2\right)\\
\Rightarrow\quad&\sum_{j=1}^{J}\|\tilde{\mathbf{A}}_j\boldsymbol{x}^{t+1}-\tilde{\mathbf{b}}_j\|^2\leq\frac{2\sum_{j=1}^{J}\left(z_0^{t+1}\|\tilde{\mathbf{A}}_j\|^2+\|\tilde{\mathbf{A}}_j\bar{\boldsymbol{x}}^{t+1}-\tilde{\mathbf{b}}_j\|^2\right)}{1-2z_1^{t+1}\sum_{j=1}^{J}\|\tilde{\mathbf{A}}_j\|^2},
\end{aligned}\tag{D.16}$$

where the last inequality is valid when z_1^{t+1} is small enough, i.e., $\{\alpha^t\}_t$ is small enough, such that the denominator is positive. Therefore, we conclude that

$$\begin{aligned}
&\sum_{j=1}^{J}I_{d,j}^{t}\left[\langle\tilde{\mathbf{A}}_j\boldsymbol{x}^{t+1}-\tilde{\mathbf{b}}_j,\boldsymbol{\lambda}_j^{t+1}-\boldsymbol{\lambda}_j^{t}\rangle-2\langle\tilde{\mathbf{A}}_j\bar{\boldsymbol{x}}^{t+1}-\tilde{\mathbf{b}}_j,\boldsymbol{\lambda}_j^{t+1}-\boldsymbol{\lambda}_j^{t}\rangle\right]\\
\leq&\sum_{j=1}^{J}I_{d,j}^{t}\left[\|\tilde{\mathbf{A}}_j\|^2\left(z_0^{t+1}+z_1^{t+1}\frac{2\sum_{m=1}^{J}\left(z_0^{t+1}\|\tilde{\mathbf{A}}_m\|^2+\|\tilde{\mathbf{A}}_m\bar{\boldsymbol{x}}^{t+1}-\tilde{\mathbf{b}}_m\|^2\right)}{1-2z_1^{t+1}\sum_{m=1}^{J}\|\tilde{\mathbf{A}}_m\|^2}\right)\right.\\
&\qquad\left.-\|\tilde{\mathbf{A}}_j\bar{\boldsymbol{x}}^{t+1}-\tilde{\mathbf{b}}_j\|^2\right]\\
\leq&\underbrace{\sum_{j=1}^{J}\left[I_{d,j}^{t}\alpha^t\|\tilde{\mathbf{A}}_j\|^2\left(1+\frac{2z_1^{t+1}\sum_{m=1}^{J}\|\tilde{\mathbf{A}}_m\|^2}{1-2z_1^{t+1}\sum_{m=1}^{J}\|\tilde{\mathbf{A}}_m\|^2}\right)\right]}_{\delta_0^t}z_0^{t+1}\\
&\qquad+\underbrace{\alpha^t\left(-1+\frac{2z_1^{t+1}\sum_{m=1}^{J}\|\tilde{\mathbf{A}}_m\|^2}{1-2z_1^{t+1}\sum_{m=1}^{J}\|\tilde{\mathbf{A}}_m\|^2}\right)}_{\delta_1^t}\sum_{j=1}^{J}I_{d,j}^{t}\|\tilde{\mathbf{A}}_j\bar{\boldsymbol{x}}^{t+1}-\tilde{\mathbf{b}}_j\|^2.
\end{aligned}\tag{D.17}$$

Note that $\lim_{\alpha^t\to 0}\delta_0^t=0$. We are now ready to show the convergence property of the BSUM-M algorithm with essentially cyclic rule. Specifically, the upper bounded of the

potential function can be expressed as follows

$$\begin{aligned}0 \leq & \sum_{t=1}^{T}\left[\triangle_p^{t+1}+\triangle_d^{t+1}-(\triangle_p^t+\triangle_d^t)\right]\\ \leq & \delta_0^1\tilde{\tau}^2\sigma\sum_{i=0}^{K}\|\boldsymbol{x}_i^0-\boldsymbol{x}_i^1\|^2+\sum_{t=1}^{T}\left(\sum_{i=0}^{K}\theta_i^t I_{p,i}^t\|\boldsymbol{x}_i^t-\boldsymbol{x}_i^{t+1}\|^2+\delta_1^t\sum_{j=1}^{J}I_{d,j}^t\|\tilde{\mathbf{A}}_j\bar{\boldsymbol{x}}^{t+1}-\tilde{\mathbf{b}}_j\|^2\right)\end{aligned} \tag{D.18}$$

where $\theta_i^t \triangleq -\frac{2\gamma+\rho\lambda_{\min}(\sum_{j=1}^{J}\tilde{\mathbf{A}}_{ij}^T\tilde{\mathbf{A}}_{ij})}{2}+(\delta_0^t+\delta_0^{t+1})\tilde{\tau}^2\sigma$. With the essentiality of the update rule, every block of variables, i.e., $\boldsymbol{x}_i$, $i=0\sim K$ and $\boldsymbol{\lambda}_j$, $j=1\sim J$, will be updated infinite times when $T\to\infty$. In the sequel, we will determine the parameter $\{\alpha^t\}_t$ for the convergence of the proposed BSUM-M algorithm with essentially cyclic rule.

The two choices of α^t, i.e., i) $\alpha^t=\alpha<<0$, $\forall\, t$, or ii) $\lim_{t\to\infty}\alpha^t=0$ and $\sum_{t=1}^{\infty}\alpha^t=\infty$, imply that for any give $\epsilon>0$, there exist an index t_0 such that for all $t>t_0$, $0<(\delta_0^t+\delta_0^{t+1})\tilde{\tau}^2\sigma<\epsilon$. Using this fact, the coefficients θ_i^t, $i=0\sim K$ and $\forall\, t>t_0$, become negative when γ satisfies (4.14). Similarly, there exists an constant t_1 such that the coefficients δ_1^t, $\forall\, t>t_1$, becomes negative. Therefore, the convergence result for the choice of α^t can be established as follows. By applying the convergence theorem of non-negative almost supermartingale [146, Theorem 1], we conclude that

$$\lim_{t\to\infty}\triangle_p^{t+1}+\triangle_d^{t+1}\text{ exists and is finite,} \tag{D.19a}$$

$$\lim_{t\to\infty}\|\boldsymbol{x}^{t+1}-\boldsymbol{x}^t\|^2=0, \tag{D.19b}$$

$$\lim_{t\to\infty}\alpha^t\|\tilde{\mathbf{A}}_j\bar{\boldsymbol{x}}^{t+1}-\tilde{\mathbf{b}}_j\|^2=0,\ \forall\, j=1\sim J. \tag{D.19c}$$

Using similar argument as in case 2 of [108, Theorem 2.1] with the diminishing stepsize rule (4.14), we can show that $\lim_{t\to\infty}\|\tilde{\mathbf{A}}_j\bar{\boldsymbol{x}}^{t+1}-\tilde{\mathbf{b}}_j\|^2=0$, $j=1\sim J$, as well. This indicates that $\lim_{t\to\infty}\|\nabla d(\boldsymbol{\lambda}^{t+1})\|=0$ holds. By (D.16), we can further show that,

$$\lim_{t\to\infty}\|\tilde{\mathbf{A}}_j\boldsymbol{x}^{t+1}-\mathbf{b}_j\|^2=0,\ \forall\, j=1\sim J\ \Rightarrow\ \lim_{t\to\infty}\|\boldsymbol{\lambda}^{t+1}-\boldsymbol{\lambda}^t\|^2=0. \tag{D.20}$$

Therefore, every limit point of $\boldsymbol{\lambda}^t$ generated by the proposed algorithm is a dual optimal solution. On the other hand, by (D.15)

$$\lim_{t\to\infty}\|\boldsymbol{x}^{t+1}-\bar{\boldsymbol{x}}^{t+1}\|^2\leq 0. \tag{D.21}$$

Hence, every limit point of $\boldsymbol{x}^t$ is a primal optimal solution.